Protein solution thermodynamics: a quasichemical perspective of solvent effects

by

Dheeraj Singh Tomar

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

July, 2014

Abstract

The unfolded(U)⇌folded(F) transition of a polypeptide chain is only marginally stable, with the net free energy change favoring F being typically about 5 kcal/mol (≈ the strength of couple of hydrogen bonds). This small stability arises from a delicate balance of large compensating contributions from hydration effects and intramolecular interactions within the polypeptide chain. Understanding this balance is of principal interest in understanding the molecular basis of forces stabilizing the protein and computer simulations can, in principle, aid in this quest.

Computer simulations have advanced to a stage where *in silico* folding of small polypeptides is now feasible, yet characterizing the hydration thermodynamics of polypeptides larger than a few amino acids remains a daunting challenge. Indeed, over the long history of molecular simulations there have been no calculations, till recently, of the hydration thermodynamics of proteins within an all-atom description of the solvent.

Building on the regularization approach to free energy calculations developed in our group, we solve the problem of calculating the hydration thermodynamics of proteins in all-atom simulations. These calculations are at a level of resolution that rival what is normally possible for simple solutes such as methane. Our framework also has the virtue of directly quantifying the hydrophobic and hydrophilic contributions to hydration, contributions that are of fundamental interest in understanding the thermodynamics of protein folding. Using the regularization approach, we have studied

the coil to helix transition of model deca-peptides and the response of this transition to thermal and chemical stresses.

A major finding of our analysis of the temperature dependent hydration of the peptides is that signatures attributed to hydrophobicity have a hydrophilic basis instead. Interestingly, these hydrophilic effects are dominated by the hydration of the peptide-backbone. Further, the textbook picture of rationalizing hydrophobic hydration in terms of specific water structures is shown to be implausible. On balance, in our model systems intramolecular interactions are as important, if not more important, than hydration effects in both the primary-to-secondary and secondary-to-tertiary structure transition.

Examining the role of chemical stresses also provides surprises, while also broadly supporting the importance of hydrophilic effects indicated by the temperature dependent hydration behavior. The denaturants urea and GdHCl are found to strengthen the hydrophobic effect for the primitive hydrophobe, a cavity that repels water, and yet they denature proteins. Within the helix-coil model studied here, urea stabilizes the coil over the helix by promiscuous hydrophilic interactions primarily mediated by dispersion forces. GdHCl, on the other hand, unfolds the helix by destabilizing it more than the coil, again by tilting the balance of hydrophilic contributions in favor of the coil. TMAO alleviates primitive hydrophobic effects and yet it drives the coil-to-helix transition, primarily by weakening the favorable hydrophilic hydration of the peptide backbone.

Tracing the reasons for the current accepted dogmas on dominant forces leads us to consider group-additive models, and these turn out to be *Achilles' heel* in modeling protein hydration thermodynamics: such models are fundamentally flawed in ignoring solvent-mediated correlations between different residues on a polypeptide. These correlations can lead to ascribing greater or lesser hydrophobicity/hydrophilicity to the defined chemical group depending on the molecular context in which the group is placed. We suggest abundant caution in relying on such group-additive approaches

ABSTRACT

for a many-body system with energy scales that can span many orders of magnitude.

This thesis, together with accumulating experimental evidence, calls for a rigorous reconsideration of the currently accepted dogmas regarding "dominant" forces driving protein folding and hence also the mechanism of folding based on such dogmas.

Thesis Adviser: Professor Dilip Asthagiri (JHU ChemBE)

Thesis Committee:
Professor Harris Silverstone (JHU Chemistry)
Professor George Rose (JHU Biophysics)
Professor Joelle Frechette (JHU ChemBE)
Professor Michael Bevan (JHU ChemBE)
Professor Jeffery Gray (JHU ChemBE)

Acknowledgments

I am grateful to:

Professor Dilip Asthagiri for mentoring me conscientiously, for mitigating the many nonacademic woes painstakingly, for leading the lab scrupulously, and for discussing science with me at odd hours and places – passionately;

Dr. Safir Merchant and Dr. Purushottam Dixit for grooming me into a graduate student, for continued suggestions on science, and for trustworthy friendships;

Dr. Valery Weber for collaborating with me, for liberally sharing computer scripts with me to make my progress faster;

Professor B. Montgomery Pettitt for collaborating with me on scientific projects, and for suggesting paths forward;

Professors George Rose for critiquing and encouraging my work, for suggesting future directions, for endorsing me, and for serving on my GBO committee;

Professor Harris Silverstone for teaching a wonderful course on quantum chemistry, for appreciating theoretical research, for always encouraging young researchers, and for chairing my GBO committee;

ACKNOWLEDGMENTS

Professors Joelle Frechette and Michael Bevan for serving on my GBO committee, and for navigating the bureaucratic hurdles in securing teaching assistantships to raise funds;

Professor German Drazer for engaging with the administration to make it gentler to students;

Professors Michael Paulaitis, Lawrence Pratt, Christopher Jarzynski, A. Ben-Naim, Robert Baldwin for critiquing my work;

Danielle Rigby for judicious accounting of funds;

Caroline Qualls for helping with the paperwork;

National Energy Research Scientific Computing Center for awarding us computer time;

ACS Petroleum Research Fund for financial assistance;

Komal and Shivendra Pandey for frequently hosting me at their place, for all the enjoyable conversations, and for lifelong friendship;

Lye Lin Lock for having the most cups of coffee with me in last four years, for regularly talking to me about our common favorite – *spicy homemade Indian tea*, and for routinely buying me groceries;

Sarat Ram, Sarita Koride, Amit Kumar, Sumedh Risbud, Rohini Gupta, Anna Coughlon, Xin Yi Chan for joining me in my hobby – eating out;

ACKNOWLEDGMENTS

And, most importantly, I am forever indebted to my family for suffering due to my absence from home during the course of my education. I dedicate this thesis to them.

We must know – we will know

David Hilbert

Contents

List of Tables

List of Figures

1

Introduction

Proteins are linear polymers of amino acids and are the hardware expression of the genetic information embedded in nucleic acids. These molecules have been termed "Nature's robots" [1] and they animate life as we know it. For example, in higher organisms, proteins are essential in fending off infections; in forming the elastic filaments that underlie the action of muscles; in storing and transporting oxygen that is then used in catabolic reactions, itself mediated by another set of proteins, to convert biochemical energy (in organic nutrients) to the energy currency ATP; in capturing photons to aid vision; ... Indeed, proteins are essential to all aspects of life [2–4].

To function properly, a protein must fold into a well-defined three dimensional structure. (Quite remarkably, a class of polypeptides exists that remains natively unfolded, but in response to environmental signals even these molecules acquire a specific conformation to affect a given task [5]. We shall not be concerned with such polypeptides in this work.) Proteins can misfold, but cellular systems appear to have mechanisms to correct for this; importantly, it is now well-appreciated that misfolding and aggregation of proteins itself causes many diseases [6]. Therefore, folding of a protein into a specific structure is an indispensable process for living organisms.

An early breakthrough in understanding the folding of the polypeptide chain came from Anfinsen's pioneering experiments on the renaturation of fully denatured ribonu-

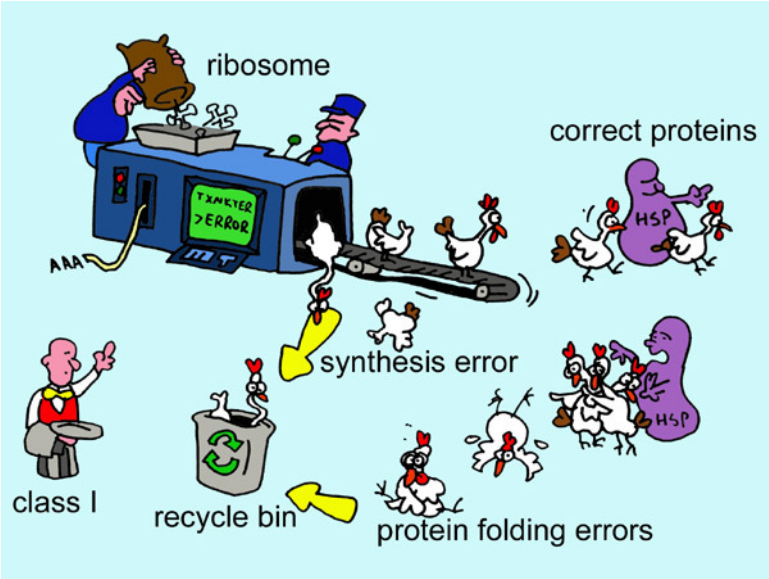


Figure 1.1: Proteins are produced in a cell using the genetic code. Healthy cells have mechanisms to recognize and destroy misfolded proteins. Picture taken from reference 7

clease. These investigations lead to the thermodynamic hypothesis that in essence implies that the unfolding $\rightleftharpoons$ folding transition is an equilibrium phenomena: " ... the three dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is one in which the Gibbs free energy of the whole system is lowest; that is, the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment." [8] The crux of Anfinsen's work is the recognition that protein folding can be treated by arguments of equilibrium thermodynamics. Anfinsen's pivotal insight was the observation that totality of interactions necessarily depends on the sequence and hence the structure, too, necessarily depends on the sequence.

Inspired by these thermodynamic arguments, Levinthal became interested in the conformational search problem – proteins fold within milliseconds despite a large

number of possible conformations. With his thought experiment, he too, like Anfinsen, highlighted the importance of the totality of these interactions [9]. In his own words, " ... protein folding is speeded and guided by the rapid formation of local interactions, which then determine the further folding of the peptide" (Ben-Naim citing Levinthal [9]). We shall not be concerned with the kinetics of protein folding in this work.

How these interactions speed and guide the folding process (their role in the structure sequence relationship) is, in fact, often called the second half of the genetic code [10]. This "code", if known, will be of immense use. For example, knowing this language can help speed the design of therapeutics to target specific proteins, in developing effective ways to crystallize and/or store protein-based therapeutics, in designing novel proteins with user-defined function, in designing other polymeric systems that could be governed by similar principles ... the potential list of scientific and technological possibilities appears limitless. It is perhaps for this reason that efforts to understand protein folding have occupied such a central place in all of biology, and biophysics.

1.1 The protein folding problem

Since Anfinsen's insight, a broad quest in protein biophysics has been to predict the structure from the sequence. This has led to what has been termed the protein folding problem, a term that oftentimes subsumes distinct facets of efforts to understand the unfolding $\rightleftharpoons$ folding transition. These facets include: (a) understanding how the sequence dictates the native state; (b) understanding how a protein finds its native structure so quickly if the number of possible structures, for a given sequence, is huge; (c) devising ways to predict the structure given the sequence.

In all facets of this problem, the solvent plays an important role, but its effect on the unfolding $\rightleftharpoons$ folding has often been estimated using rather strong assumptions

and/or approximations. Reassessing the role of the solvent using a first-of-its-kind framework that allows one to avoid making the strong assumptions and/or approximations made earlier is a principal achievement of this work. Our approach also calls for a reconsideration of the current accepted ideas about dominant forces in protein folding, and thus potentially also the mechanism of protein folding.

We next briefly review the extant ideas on the dominant thermodynamic forces in folding and its role in the folding mechanism. (We borrow liberally from Tanford's book [1] summarizing historic developments in understanding proteins and protein folding.) After this brief (but not exhaustive) summary we will present a broad overview of our approach and contributions.

"Thermodynamic forces" popularly used to explain protein folding can be separated in two broad but distinct categories: hydrophobic forces and hydrophilic forces. Normally perceived, these "thermodynamic forces" are exclusive as they emerge from a diverse set of atomic groups (see Figure 1.2 for surface diversity) within a protein.

1.1.1 Hydrophobic effects

The hydrophobic effect is a widely invoked "thermodynamic force" to explain protein folding. The similarities amongst protein folding, oil/water demixing, micelle formation, and surfactant adsorption on water surface have motivated researchers to define a "thermodynamic force" that is common to all these phenomena.

The following properties give an impression of hydrophobic behavior: (a) solubility so low that Henry's law becomes applicable (positive excess chemical potential); (b) further decrease in solubility with increase in temperature implying negative excess entropy of hydration; (c) high and positive excess heat capacity of hydration (heat capacity of hydration of hydrophobic solute is significantly more positive than the heat capacity of the same solute in its pure form [11]); (d) clustering of hydrophobic molecules in water (negative osmotic second virial coefficient). However, some solutes

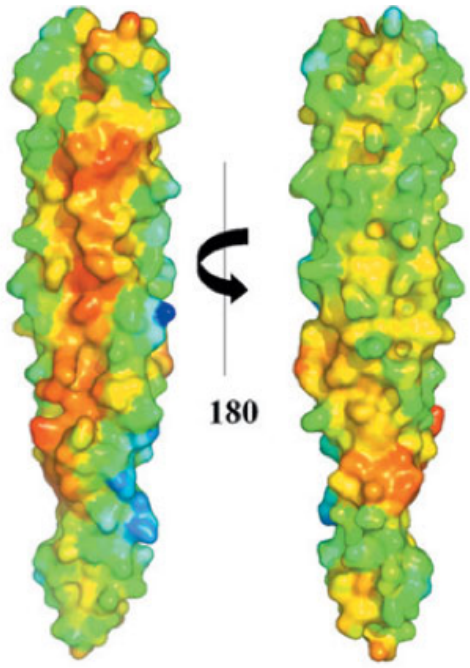


Figure 1.2: A representative picture of the diversity of amino acids present in a protein. Bigger, in size, hydrophobic groups (red), smaller hydrophobic groups (green and yellow) along with ionic residues (blue) are coated on a long chain of backbone that can make hydrogen bonds with the solvent. Picture is only for the purpose of demonstration and taken and modified from reference 12

that do not satisfy all four properties are still considered hydrophobic. For example, Tetrahydrofuran is a water soluble compound, yet it is considered hydrophobic due to (b) and (c), as no actual evidence of clustering of Tetraydrofuran molecules exists. Nevertheless, these four properties define the uniqueness of oil and water demixing and thus hydrophobic behavior.

The credit for recognizing the importance of hydrophobic effects goes to Isidor Traube, who, in 1891, showed that the lowering of surface tension of a homologous series of compounds depends on the length of the nonpolar chain [13]. The clear inference was that the nonpolar tails preferentially partition to the air/water interface

rather than be solvated in water. Similar ideas also led Danielli [14] to suggest that proteins might fold to hide their nonpolar pieces from water. Later Langmuir, who had previously used his understanding of hydrophobicity to explain soap micelle formation, together with Wrinch [15], proposed an important role of hydrophobicity in the protein folding. Contemporaneous studies on the solubility of nonpolar gases in liquid water already revealed peculiar entropic and heat capacity signatures. Frank and Evans [16] later rationalized these signatures by introducing the iceberg model. This model hypothesizes that water forms cages or clathrate-like structures around the small apolar solutes; in this model, the formation of the cage is used to rationalize the negative entropy of hydration and the latent heat required to melt the cage with the high and positive heat capacity of hydration. It was within this intellectual milieu that Kauzmann [17] placed the hydrophobic effect at the center of the protein folding problem and likened hydrophobic interactions to hydrophobic bonds that were formed by the drive to librate water from the cages surrounding the nonpolar solute. (As we shall note below, perhaps the then evolving understanding of the role of hydrogen bonding in stability also biased Kauzmann's view, but one that he continued to champion till later in his life [18].) This picture going back to Frank-Evans and Kauzmann is the canonical explanation found in textbooks to explain hydrophobic behaviour and oil-water demixing.

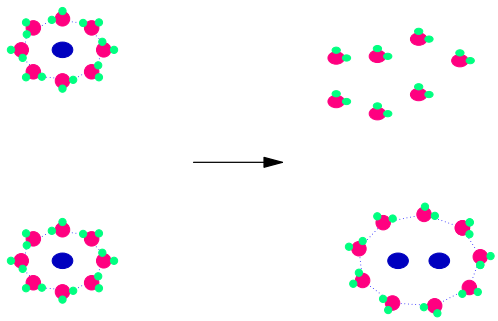


Figure 1.3: A demonstration of hydrophobic bond. Hydrophobic molecules (blue) come closer to each other and the water molecules liberate from the cages

Later, Tanford used the then available data (with some approximations regarding group additivity and chain entropy) and argued that the need to sequester hydrophobic residues from water can easily compensate for chain entropy that would drive the polymer to unfold [19]; this further strengthened the case for hydrophobic hydration in folding.

These works on the clustering of hydrophobic molecules appear to have led to a pleasing and simple picture, the protein folding funnel [20]. In this model, Anfinsen's restatement of the second law of thermodynamics for proteins in a solvent is, often contentiously [9, 21], interpreted as "free energy minimum of **a** protein" in a solvent instead of "free energy minimum of the protein **and** the solvent" – a subtle yet profound change. (We note, especially in attempts to deduce structure from sequence, the role of solvent is usually relegated to a featureless continuum, and the discussion focuses on the free energy minimum of the protein [22].) That the "protein folding landscape" is shaped like a funnel and unfolded structures roll downhill (or collapse) to stable native structures at the bottom of the funnel is explained by the drive of hydrophobic groups to sequester in the interior of a protein, much like how soap micelles form in water.

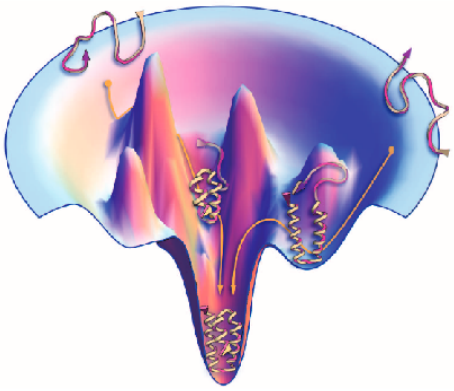


Figure 1.4: The folding funnel model. "Protein folding landscape" is shaped like funnel. Figure taken from reference 20

Now, the hydrophobic effect and the folding funnel model are ubiquitously present in biophysics textbooks and they are taught to students as standard explanations for protein folding [23–30].

An important contribution of this thesis is to show that the structural interpretation of hydrophobic effects is not correct, and in fact the signatures of hydrophobic effects can arise due to hydrophilic effects. Thus the very idea of oil-water demixing as being relevant to the polypeptide chain collapse must be reconsidered.

1.1.2 Hydrophilic effects

Attractive interactions provide the cohesion that holds any material together, and their role can be most readily appreciated in phase transitions such as melting and boiling, non-ideal behavior of gases, non-ideal behavior of salts in solutions, etc. However, their perceived role in the unfolding $\rightleftharpoons$ folding transition of proteins has waxed and waned. (Hydrophobic effects also provide an effective attractive force between two nonpolar solutes, but the basis of this force, as noted above, is in the behavior of the solvent. Hydrophobic bonds thus have an entropic origin, whereas by hydrophilic effects, we emphasize the enthalpic aspect of interactions.)

The discussion of attractive interactions is facilitated, and the points (to be made later in the thesis) further sharpened, by considering attractive interactions at different length-scales: (1) Long-range electrostatic interaction between charged groups and between the charged groups and the solvent. (By long-range it is understood that the residues are typically separated such that there is no possibility of short-range hydrogen-bonding.) (2) Short-range hydrogen bonding interactions. In both cases, dispersion interactions can play an important role, but this aspect has received surprisingly little detailed treatment.

In protein solution thermodynamics, electrostatic interactions influence pH dependent behavior, pH-regulation of enzyme activity, acid and alkaline denaturation,

and protein substrate/inhibitor interactions [31]. The earliest efforts to understand such effects goes to Linderström-Lang and his study of interaction between charged residues on the basis of the the then newly developed Debye-Hückel theory [31]. Within this classical picture, since the electrostatic energy depends on the square of the net charge, a view appears to have emerged that "no electrostatic contribution to protein stability is expected near the isoelectric point" [32]. Thus long-range electrostatics have typically been discounted in attempts to understand the dominant forces in the unfolding $\rightleftharpoons$ folding transition of proteins. But a protein with no formal net charge still has partial charges and the interaction of these partial charges among themselves and with the solvent can still contribute to stability, an aspect that, quite surprisingly, remains underappreciated. (We return to this issue later in highlighting the role of such interactions in the coil to helix transition.)

Turning now to short-range attractive interactions, in early attempts to explain protein denaturation (when the three dimensional structures for proteins were yet unknown), Mirsky and Pauling insightfully anticipated the role of attractive interactions involving >NH$\cdots$O=C< pairs in stabilizing proteins. (These groups, of course, constitute parts of the polypeptide backbone.) Pauling, together with Corey and Branson, later ingeniously used the need to satisfy such hydrogen bonds to propose two new structures of polypeptide chains, the alpha-helix and the beta-sheet, structures that were later confirmed by x-ray scattering. [2, 4]

The suggested importance of hydrogen-bonding likely spurred the quest to quantify its role in stabilizing proteins. Using urea-urea pairing as a model for hydrogen bond formation, Schellmann concluded that hydrogen bonds only provide a marginal stability to proteins. (See Kauzmann's review [17] and citation to Schellmann's work; as noted in the earlier section, this suggested weakness of h-bonding was another likely reason in Kauzmann's emphasis on the hydrophobic effect.) Klotz and Franzen [33] came to a similar conclusion on the basis of free energy of pairing between N-methylacetamide molecules, and somewhat later, the hydrogen bond inven-

tory argument (see equation 1.1, net gain in number of hydrogen bonds is zero after unfolding) by Fersht [34, 35] also led to the notion that hydrogen bonds only provide marginal (to non-existent) stability to proteins.

$$\begin{array}{ccccc} & \text{Folded state of a protein} & & \text{Unfolded state of a protein} & \\ >\text{NH}\cdots\text{O}=\text{C}< & + & \text{water}\cdots\text{water} & \rightleftharpoons & >\text{NH}\cdots\text{water} + \text{water}\cdots\text{O}=\text{C}< \\ \text{Number of hydrogen bonds} = 2 & & & & \text{Number of hydrogen bonds} = 2 \end{array} \tag{1.1}$$

The efforts by Schellmann, Koltz and Franzen, and Fersht all rely on pairing between model compounds that are free to roam in a solvent to infer properties of such bonds between groups that are otherwise anchored on a polypeptide chain. Besides this, as Ben-Naim has highlighted [36], arguments such as the h-bond inventory (and by extension similar counting arguments extant in the literature), are fundamentally faulty in mixing hydration Gibbs energy with solute-solvent binding energy and neglecting the fact that waters of hydration, even when they are released from the solute, still remain at the same chemical potential (and hence their pairing or not cannot contribute to a driving force as indicated in the stoichiometric equation 1.1)

Later, a decade after presenting the argument, Fersht himself became critical of the oversimplification: "Although the hydrogen bond inventory is zero, hydrogen bonding is energetically favorable in the formation of enzyme-substrate complexes because of the increase in the entropy on the release of bound water molecules .[This is analogous to] intramolecular reactions versus intermolecular reactions ... When two molecules associate, they lose translation and rotational entropy, and this has to be deducted from the energy balance on binding. But if they are already held together in complex, there is no further loss of translational and rotational entropy if an extra hydrogen bond is built in" [35].

(Theoretical and computational efforts [37–40] efforts to understand the role of

hydrogen bonding in protein stability also provide an unclear picture. We can note that even in the prototypical problem of helix to coil transition, using continuum dielectric models of the solvent, one set of calculation suggested that hydrogen bonds do not stabilize the helix [41] whereas another came to an opposite conclusion [42].)

Within the last 15 years, studies on the role of osmolytes in protein stability has once again refocused attention on the importance of the role of attractive interactions, in general, and hydrogen bonding, in particular, in protein folding [43, 44]. These studies indicate that osmolytes affect protein stability principally by their action on the peptide backbone; the inference is clear that this implies interaction of the backbone itself must be important in protein folding. During about the same period, Ben-Naim has helpfully emphasized the importance of all hydrophilic effects in protein folding [45]. Further, experiments within the last few years have shown that, quite remarkably, backbone itself can collapse (no oily residues necessary!) and fold a protein [46].

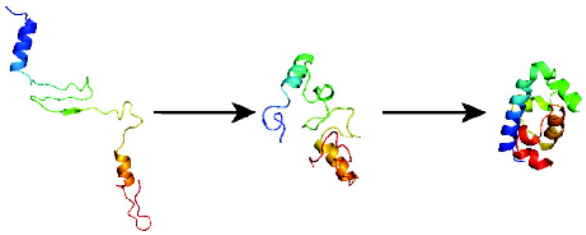


Figure 1.5: A picture to demonstrate hierarchic folding model. Folding begins in the neighborhood. Lower level of structures are formed first. These structures later form tertiary structure. Picture taken from Reference 50. Note: This reference is not related to the hierarchic assembly model

The importance of hydrophilic effects, especially the role of hydrogen bonds, naturally suggests that local interactions along the polypeptide chain should be important in folding. This suggestion finds its best expression in the hierarchic assembly model of folding. (We emphasize that hierarchic assembly does not imply non-local attrac-

tive interactions are not important [47, 48].) Indeed one of the important ideas of this model is that local nucleation events could be stabilized by non-local interactions [49]. An aspect of such an effect can already be inferred from our study on the helix/coil transition and the pairing of helices.

1.2 Overview

Experimental studies that seek a molecular rationalization of the observed thermodynamics of protein folding invariably appeal to hydration of small molecule analogs of the protein groups. In theoretical approaches to assess the thermodynamics of hydration, the studies have typically focused on idealized models that seek to better illuminate the physics under study, for example, for assessing the potential role of hydrophobicity in folding, researchers have sought to understand the collapse of idealized alkane chains. Often at the level of proteins, the solvent is usually described as a featureless continuum [51, 52]. As Ben-Naim noted in 2011, no direct assessments of the free energy of a protein in a solvent described in full atomistic detail existed (but see 2012 work from our lab [62]).

While simulations of protein folding have matured, assessing the thermodynamics of protein folding and the role of solvent in this process have been beset by one main challenge: To calculate free energy of protein in a given conformation with fidelity to the atomistic nature of the solute and the solvent while simultaneously segregating all the contributions to the free energy (viz. hydrophobic, hydrophilic, electrostatic, etc.) to uncover "thermodynamically dominant forces" in unfolding $\rightleftharpoons$ folding transition of proteins.

We have, for the first time in the history of this problem, succeeded in the effort to meet this challenge by developing a computational framework to apply the quasi-chemical theory of solutions to molecules as big as proteins. Thus it is now possible to study the hydration thermodynamics of a protein in a given conformation at the

same level of resolution that was formerly only possible for simple solutes such as methane.

We will discuss our findings for thermodynamics of unfolding $\rightleftharpoons$ folding transition of proteins using this framework in this thesis. We have organized the rest of the thesis as follows.

We will introduce quasichemical theory of solutions and the framework we have developed to implement it in chapter 2.

In chapter 3, we will study the hydration of deca-peptides in response to a thermal perturbation. In chapter 4, we will study the transfer of deca-peptides from water to an aqueous solution of osmolytes; here the protein thermodynamics is perturbed by a solvent additive. The results of both chapters lead to surprising findings. The main role of these perturbations is in modulating protein-solvent attractive interactions, rather than the hydrophobic effect. If the temperature signatures of hydrophobic effects as suggested by calorimetric experiments, are invalid for protein solvation, then why did many studies support the hydrophobicity hypothesis?

In chapter 5, we will discuss the group additivity based assumption (the fundament of many seminal studies that supported hydrophobicity hypothesis) to answer the question why the hydrophobicity based hypothesis appeared successful. We will ask two questions here: Are group additivity based calculations accurate? and, can group additivity discover mechanism of protein folding? We will demonstrate that how competing errors can cancel each other and give the illusion that group additivity accurately calculates transfer free energies. We will show that hydration free energies are not context independent and powerful near neighbor correlations, mediated by the solvent, make group additivity ill suited to analyze the mechanism of protein folding.

Having established the importance of hydrophilic effects in folding and the potential flaws in additive models, we apply our methods to explore the primary-to-secondary (coil-to-helix) and secondary-to-tertiary (helix-helix pairing) transition in a simplified, model system. (In chapter 6, we will take a deca-peptide and give rig-

orous thermodynamic treatment to its folding sans any approximations.) This work will highlight the importance of attractive interactions and build a convincing case for protein folding being a delicate balance of attractive interactions instead of domination by hydrophobic effects. Additionally, we will show here that the long-range non-specific attractive interactions play a non-trivial role in the stability of the tertiary structure of a protein.

In chapter 7, we will conclude by summarizing our perspective about the role of solvent in protein folding that we have developed after afore-mentioned studies. We will also discuss promising future applications of our framework.

2

Theory and Methods

One of the principal objects of research in any department of knowledge is to find the point of view from which the subject appears in the greatest simplicity.

J. W. Gibbs

The chemical potential of a molecule, μ_α, in a solution of volume V at inverse thermodynamic temperature $\beta(=\frac{1}{k_B T})$ is given by

$$\beta\mu_\alpha = \ln \frac{\Lambda_\alpha^3}{\mathrm{V}\ q_\alpha^{\mathrm{int}}} + \beta\mu_\alpha^{\mathrm{ex}} \tag{2.1}$$

The first quantity on the right hand side is the ideal contribution to the chemical potential and it is independent of solvent; the thermal de Broglie wavelength, Λ_α, and the internal partition function of the molecule, q_α^{int}, depend on temperature and the properties of the solute alone.

The second quantity on the right hand side, μ^{ex}, known as excess chemical potential, holds the key to understand the role of solvent in the chemical potential of the molecule, as it is related to the solute-solvent interaction energy, ϵ. For a solute α [55, 60] with the distribution of solute-solvent binding energy, $P(\epsilon)$,

$$\beta\mu_\alpha^{\mathrm{ex}} = \ln \int P(\epsilon)\ e^{\beta\epsilon} \tag{2.2}$$

Unfortunately, this form of the potential distribution theorem (PDT) is useless in computer simulations as the tail of the interaction energy distribution, $P(\epsilon)$, is not well behaved (Figure 2.1). Fortunately, this statistical problem of ill-behaved tails of interaction energy in the PDT can be regularized with the help of external constraints [56–71]. This approach leads to a form of the quasichemical organization of the PDT. We next present a brief overview of the quasichemical theory of solutions.

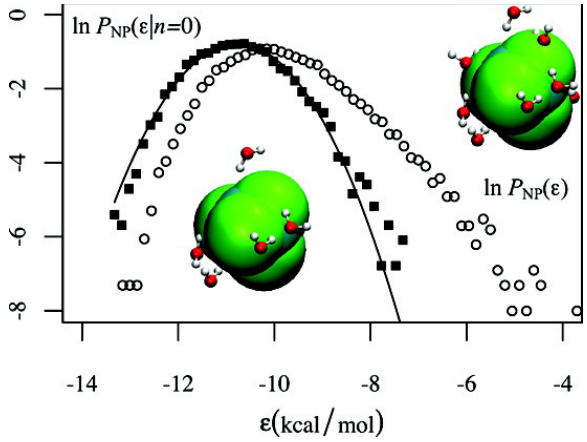


Figure 2.1: Binding energy distribution of neopentane ($NP = C(CH_3)_4$) to water. The open circles are the probability distribution without regularization. The filled squares are the regularized distribution. The distribution after regularization is a well behaved Gaussian model, a signifiant improvement over the distribution without regularization. Figure taken from reference 68

The development here is based on our previously published paper [72].

2.1 Quasichemical theory

As Figure 2.1 shows, the long-tails reflecting high solute-solvent interaction energies limits the direct use of Equation 2.2 as these high-energy tails are never well-sampled. Imagine now a field $\phi_R(r)$ centered on the solvent that serves to push the

solvent away from the solute-solvent interface. As the solvent is pushed away, we temper the high-energy solute-solvent interactions and the conditioned $P(\epsilon)$ starts approaching a Gaussian. This modified problem is now easier to solve because if $P(\epsilon)$ is a Gaussian distribution (with mean $\langle\varepsilon\rangle$ and variance $\langle\delta\varepsilon^2\rangle$) then the integral in Equation 2.2 simply reduces to $\langle\varepsilon\rangle + \frac{\beta}{2}\langle\delta\varepsilon^2\rangle$.

Now we need to account for applying the field, and this is done in both the absence and presence of the solute in the solvent. (See reference 58 for rigorous derivations.) This then leads to

$$\beta\mu^{\text{ex}} = \underbrace{\ln x_0[\phi_R]}_{\text{local chemistry}} \underbrace{-\ln p_0[\phi_R]}_{\text{packing}} + \underbrace{\beta\mu^{\text{ex}}[P(\varepsilon|\phi_R)]}_{\text{long-range}} \tag{2.3}$$

Figure 2.2 provides a pictorial representation of the above equation. Here, $-\ln x_0[\phi_R]$ is the free energy required to apply the field in the solute-solvent system: it reflects the strength of the solute interaction with the solvent in the inner shell (naturally because this quantity is obtained after pushing the solvent molecules away from the solvent). $-\ln p[\phi_R]$ is the free energy required to apply the field in the neat solvent system: it reflects the intrinsic properties of the solvent. For ϕ_R modeling a hard exclusion of solvent, $-\ln p_0[\phi_R]$ is precisely the hydrophobic contribution to hydration [56, 61]. $\beta\mu^{\text{ex}}[P(\varepsilon|\phi_R)]$ is the contribution to $\beta\mu^{\text{ex}}$ from long-range solute-solvent interactions. In molecular dynamics simulations, we calculate $-\ln x_0[\phi_R]$ or $-\ln p_0[\phi_R]$ simply by the work required to apply ϕ_R. This field ϕ_R, pushes the solvent molecules away thus creating a cavity (see Figure 2.3).

After briefly sketching the quasichemical theory to calculate chemical potential of a solute, we will now focus our attention to another quantity of interest, namely the entropy of solvation. In the next section we will briefly present an approach to calculate the entropy of hydration for an atomic solute. (The same framework, of course, applies to molecular solutes as well.)

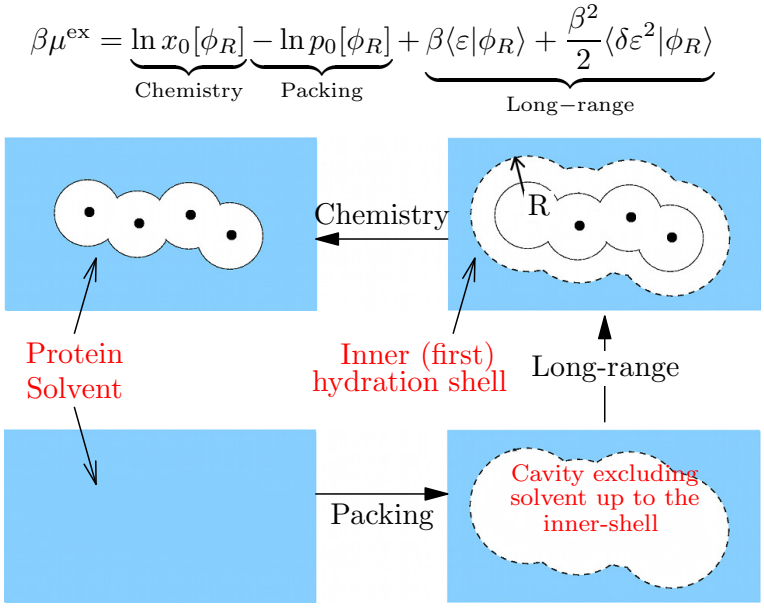


Figure 2.2: Quasichemical theory for molecular solutes. Solvation shell of a molecule is a contour formed by many spheres, each centered on a heavy atom of the solute. Figure taken and modified from reference 72. First a cavity of the size and shape of the solute molecule is created (packing) pushing away solvent molecules with the help of a field, then the solute is put inside this cavity (long-range). Observe when long-range is calculated there are no near neighbor contacts between solute and solvent. In the end the solute is allowed to make near neighbor contacts (by turning off the cavity) thus giving the strength of near neighbor interactions.

2.2 Enthalpy and entropy of solvation

The excess enthalpy (h^{ex}) and excess entropy ($s^{\text{ex}} = -\left(\frac{\partial\mu^{\text{ex}}}{\partial T}\right)_{T,P}$) of hydration of a solute are given by equations 2.4.

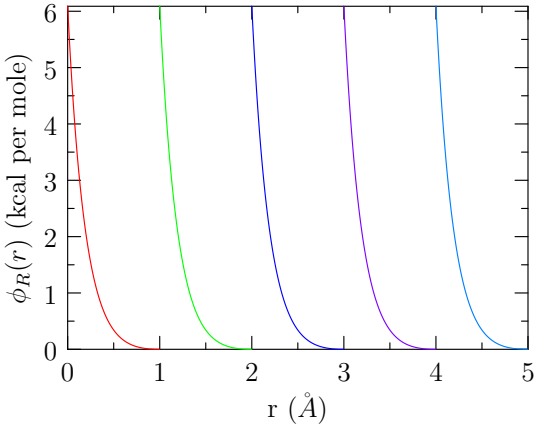
$$\begin{aligned} Ts^{\text{ex}} &\approx E_{sw} + E_{reorg} - \mu^{\text{ex}} \\ h^{\text{ex}} &\approx E^{\text{ex}} = E_{sw} + E_{reorg} \end{aligned} \tag{2.4}$$

$$\phi_R(r) = 4a\left[\left(\frac{b}{r-R+\sqrt[6]{2}b}\right)^{12} - \left(\frac{b}{r-R+\sqrt[6]{2}b}\right)^{6}\right] + a; \text{ if r} \leq \text{R}$$

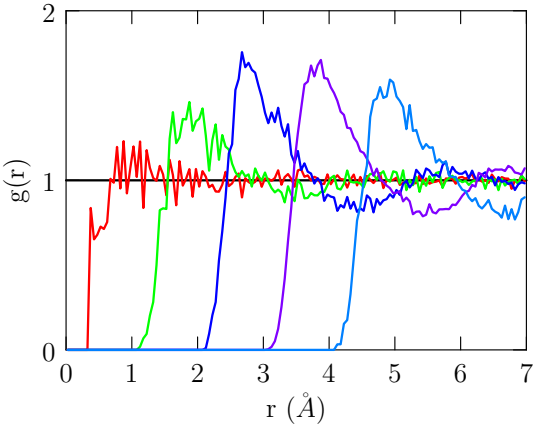
$$\phi_R(r) = 0; \text{ if r} > \text{R}$$

$$a = 0.155\ \tfrac{\text{kcal}}{\text{mol}}; \quad b = 3.1655\ \mathring{A}$$

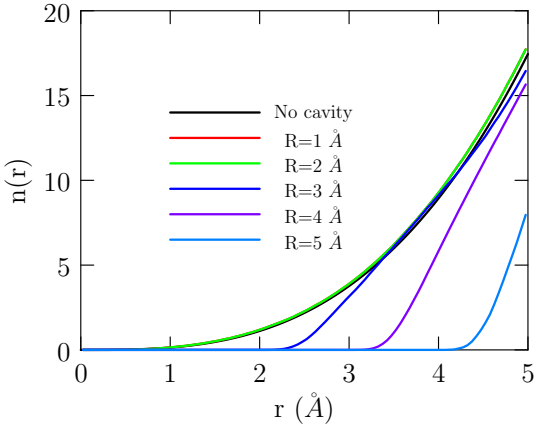
(a) Mathematical form of the field



(b) Graphical representation of the field



(c) Radial distribution function from the center of the cavity



(d) Number of water molecules in a sphere of radius r concentric with the cavity

Figure 2.3: Cavity in pure water at 298 K and 1 bar. For comparison, cavity of zero radius (no cavity) is also shown. As the cavity grows water molecules are pushed further away from the center of the cavity. For a sufficiently big cavity, tails of binding energy distribution behave nicely. All panels have same legend (here shown only in the bottom-right panel).

(For a detailed theoretical development of equations 2.4 see Appendix on page 89. In these equations we neglect tiny contributions from the finite thermal expansivity and isothermal compressibility of the solvent. We additionally ignore small contri-

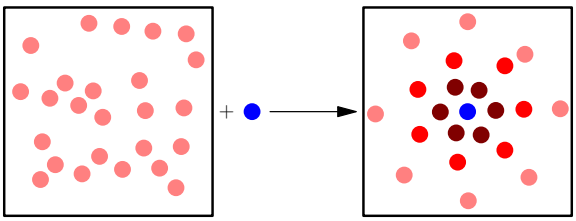


Figure 2.4: Pictorial representation of solvent reorganization. After the solute (shown in blue) is introduced to the system (right panel), solvent molecules close to the solute (shown in brown and red) have different potential energy (average of interaction energy with the rest of the molecules of the solvent) than the solvent molecules far away from the solute (shown in light red). The solvent molecules far away from the solute have the same energy as the bulk (left panel).

butions from a finite excess volume of hydration. These conditions are satisfied for molecules studied here.)

E^{ex}, is the sum of solute-solvent binding energy (E_{sw}) and reorganization energy (E_{reorg}). Reorganization energy is defined as average potential energy of the solvent in the solute-solvent system minus the average potential energy of the solvent in neat solvent system. The solvent molecule close to the solute contribute the most to the reorganization energy as their hydration shell is affected the most by the presence of the solute, but farther away from the solute-solvent interface, we expect the potential energy of a water molecule to be close to the bulk values. Note also that the solute-solvent interaction contribution, E_{sw}, can itself be decomposed due to contributions from the backbone (E_{bb}) and the sidechain (E_{sc}), in pairwise additive forcefields.

We will now present methods that are common to all our calculations used in our simulation: any specific additions to this will be noted in respective sections.

2.3 Methods

The solvent was modeled by the TIP3P [110, 112] model and the CHARMM [114] forcefield with correction terms for dihedral angles [113] was used for the peptides.

We apply atom-centered fields to carve out a molecular cavity in the liquid, in contrast to reference [62] where the external field evacuated a spherical domain around the molecule. The functional form of the field is noted in Figure 2.3. To build the field to its eventual range of $R = 5$ Å, we progressively apply the field, and for every unit Å increment in the range, we compute the work done in applying the field using Gauss-Legendre quadratures.
Five Gauss-points $\left[0, \pm(1/3)\sqrt{5 - 2\sqrt{10/7}}, \pm(1/3)\sqrt{5 + 2\sqrt{10/7}}\right]$ are chosen for each unit Å. At each Gauss-point, the system was simulated for 1 ns and initial data points were excluded for equilibration. (Excluding more data did not change the numerical value significantly, indicating good convergence. Error analysis and error propagation was performed as before [62].) The starting configuration for each R point is obtained from the ending configuration of the previous point in the chain of states. For the packing contributions, thus a total of 25 Gauss points span $R \in [0, 5]$. For the chemistry contribution, since solvent never enters $R < 2.5$ Å, we simulate $R \in [2, 5]$ for a total of 15 Gauss points. Separate calculations with a lower order Gauss-Legendre quadrature and a trapezoidal rule (with R incremented in steps of 0.1 Å [62]) showed that results are very well converged with the five-point quadrature (data not shown).

The long-range contribution $\mu_n^{\text{ex}}[P(\varepsilon|\phi_R)]$ was obtained by inserting the solute [63] in a cavity (with atom-centered radius $R = 5$ Å).

All calculations are performed at 1 bar pressure and with a time-step of two femtoseconds. Temperature is maintained at 298.15 K (unless specified otherwise). Verlet algorithm is used to integrate the equation of motion. Langevin dynamics is used to control temperature and pressure. Damping coefficient is set to 1 per pico-second. Langevin piston period is 200 femto-seconds. Langevin piston decay

is 100 femtoseconds. Rigid bonds are used for water with the SHAKE algorithm, and particle mesh Ewald technique is used to for electrostatic calculations with a grid spacing of 0.5 Å. Van der Waals interactions are smoothly turned to zero at a distance of 10.43 Å beginning at a distance of 9.43 Å.

We calculate E_{reorg} by the hydration-shell-wise procedure earlier developed for methane [71]. (See Figure 2.3 for a pictorial representation for an atomic solute; hydration shell of the peptide, as is the case here, is the collection of spheres of radius R centered on each heavy atom) We define first hydration shell as $R \leq 5$ Å (our regularization for excess chemical potential also uses the same R). Water molecules that satisfy $5.0 < R \leq 8$ Å are counted as second hydration shell and water molecules that satisfy $8.0 < R \leq 11.0$ Å are considered part of third hydration shell.

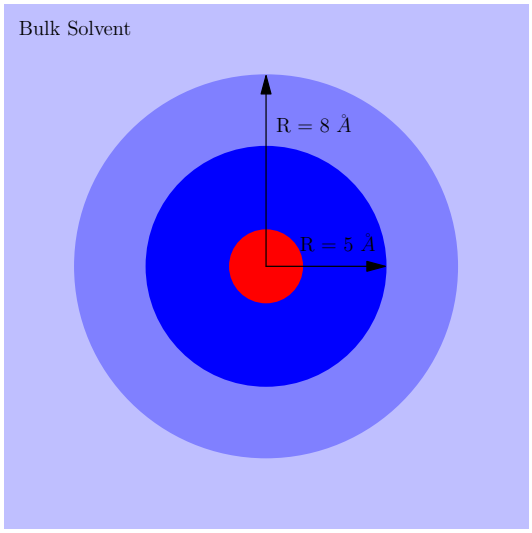


Figure 2.5: Pictorial representation of shell based method to calculate solvent reorganization energy. First shell solvent molecules (shown in blue color) are closer than 5 Å to the solute (red); second shell molecules (shown in light blue) are farther than 5 Å but closer than 8 Å. Molecules farther than 8 Å do not contribute much to the reorganization energy (shown here as bulk; lighter blue).

3

Temperature signatures of hydrophobicity are hydrophilic in origin

Note: This chapter draws upon a manuscript titled "Some surprises in the temperature dependent hydration of an oligopeptide: signatures of hydrophobic hydration have a hydrophilic basis" that is currently being circulated to other researchers to obtain their comments prior to submission.

3.1 Overview

A direct assessment of temperature dependent hydration of model proteins shows that signatures that appear to suggest hydrophobic hydration on the basis of similar signatures in small molecules (discussed in Chapter 1) have a hydrophilic basis instead. Interpreting the negative entropy of hydration in terms of specific water structures is also shown to be deficient. Thus translating notions of hydrophobicity from small nonpolar molecules to model similar physics in proteins, an approach that has been instrumental in the genesis of extant views on the role of hydrophobicity in

protein folding, should be reconsidered.

3.2 Introduction

Oil and water do not mix under conditions of temperature and pressure that are of most interest in biology. This phobia of non-polar solutes to dissolve in water, the hydrophobic effect, has been suggested as an important [17], perhaps even the dominant [18, 32, 146], thermodynamic driving force that causes an unfolded polypeptide to fold into a globular structure with the majority of the nonpolar residues hidden in the interior of the structure, away from the solvent [18, 32, 147].

Hydrophobic hydration of non-polar gases [11] is characterized by two distinct temperature signatures: (1) the partial molar excess entropy of hydration is negative. (Throughout this work excess quantities are defined for transfer from vapor to liquid with solute density the same in either phase; the partial molar entropy implied here is also the unitary or contact entropy noted by Kauzmann [17].); and (2) the corresponding heat capacity increment upon hydration, the partial molar excess heat capacity, is positive. The partial molar enthalpy of hydration is negative (favoring hydration), but, on balance, the partial molar Gibbs free energy of hydration is *positive*. These temperature signatures are also observed for the transfer of a non-polar solute from an apolar liquid to water, provided care is used in accounting for changes in dilution upon hydration [17, 103].

Following Frank and Evans [16], the entropy and heat capacity signatures have been usually rationalized by invoking some form of ordering in the layer of water surrounding the solute [18]. This rationalization is intuitively appealing since induced order suggests a negative entropy of hydration and the greater amount of thermal energy to melt the ordered layer of water agrees with the positive heat capacity increment on hydration. This is also the model of hydrophobic hydration that is either implied or explicitly stated in modern textbooks [23–30]. .

3. SIGNATURES OF HYDROPHOBICITY

The hydration thermodynamics of small non-polar solutes and its structural rationalization suggests that pairing between two hydrophobes will be favored by entropy gained upon release of water molecules. Drawing a parallel between this suggestion, the substantial prevalence of amino acids with non-polar sidechains (and the phobia of non-polar solutes to mingle with water), and available data on thermodynamics of protein aggregation, hydrophobic hydration was originally suggested to be an important factor in protein folding [17]. (Around the same time the stabilizing role of hydrogen bonds also came to be questioned (for example, reference 33), perhaps further strengthening the case for hydrophobic hydration as a key factor in protein stability.) A subsequent quantitative analysis of hydrophobic hydration in protein unfolding [19, 148], albeit one using a group additive decomposition of the hydration free energy into contributions solely from the non-polar parts of amino acid side chains, suggested that the unfavorable (hydrophobic) hydration of the non-polar parts easily balances the chain entropy gained upon unfolding, further supporting the importance of hydrophobicity in folding.

Microcalorimetry has provided the most direct experimental information on the temperature dependence of protein unfolding [95, 102] and hence on the possible role of hydrophobicity. These studies clearly showed a positive heat capacity change upon protein unfolding, a feature that was rationalized as arising mainly due to the heat capacity increment from the exposure of hydrophobic residues upon protein unfolding [102]. (The then available heat capacity data also correlated with the number of nonpolar contacts in the proteins studied.) Importantly, for some of the proteins, the temperature-dependent specific enthalpy and specific entropy of unfolding also appeared to converge to a single value at a temperature around 110 °C [95, 102]. This feature hinted at a deeper regularity holding the protein structure together; in conjunction with the heat capacity data, the importance of hydrophobicity again appeared to be confirmed. However, later more extensive studies have questioned the reliability of the convergence behavior [149]. A later analysis of the heat capacity

data on the basis of group-additivity suggested that hydration of polar and aromatic groups favor protein unfolding and overwhelm the stabilizing influence of hydrophobicity [116], but the validity of this inference is in doubt in light of the emerging understanding of limitations of group-additivity (chapter 5).

Theoretical studies on the hydration of a cavity, an ideal or primitive hydrophobe, have been instrumental in furthering the current understanding of hydrophobic hydration [56, 61, 127–129]. The hydration of such cavities naturally arises in our approach to calculate the excess free energy. Experimentally the hydration of nonpolar solutes is used to model hydrophobic hydration, but as we shall see below the hydration of such solutes also requires one to consider solute-solvent attractive interactions, effects that are properly classified as hydrophilic aspects of hydration.

The hydration of cavities crucially highlights the length-scale dependence of hydrophobic hydration [127–129]: the hydration thermodynamics of small solutes scales with the solute volume, whereas for larger solutes it depends on the solute-solvent interfacial area. (For our purposes, side-chain analogs for nonpolar sidechains belong to small-solute category, polypeptides and globular proteins belong to the larger solute category.) This observation raises the question whether translating small solute-hydrophobicity to model hydrophobicity at the level of proteins using a group-additive scheme is valid, a concern that is amplified by recent studies that question additivity [72, 84, 122, 123]. Moreover, the temperature dependences of the small and large length-scale primitive hydrophobic hydration are also expected to be different. For small (2-4 Å radii) solutes, simulations [129] show (as is also seen in this work) that the excess free energy increases with temperature for biologically-relevant temperatures, whereas the converse is expected at the large scale regime described by the free energy to create a vapor-liquid [150, 151] interface. Thus it remains unclear whether heat-capacity increments seen in microcalorimetry can be attributed to hydrophobic hydration of the newly exposed nonpolar groups.

Earlier theoretical work has greatly advanced our present understanding of hy-

drophobic hydration, but even for the prototypical problem of coil-to-helix transition, a detailed examination of the temperature-dependent hydration thermodynamics is as yet unavailable. Here we use our approach to free energy calculations to study temperature-dependent hydration effects in the classical problem of the coil-to-helix transition in a deca-alanine peptide. Our work turns up a surprise: signatures that are attributed to hydrophobicity arise instead from hydrophilic hydration, primarily of the peptide-backbone.

Together with the the observation that oligoglycine can itself collapse [46, 53, 54], experimental studies indicating the role of backbone-driven collapse of polypeptides [46] and osmolyte-driven contraction of unfolded proteins [93, 152], our work then suggests that we must reconsider the extant hydrophobic-collapse-based model of protein folding, and indeed the current understanding of dominant forces in protein folding.

3.3 Methods

In addition to the methods listed in section 2.3 on page 21, we follow the following methods.

We model the helix and coil states of a deca-alanine peptide. The peptides were modeled with an acetylated (ACE) N-terminus and n-methyl-amide (NME) capped C-terminus. The extended β-conformation ($\phi, \psi = -154 \pm 12, 149 \pm 9$) was aligned such that the end-to-end vector lay along the diagonal of the simulation cell; the helix was aligned with its axis along the x-axis of the cell. The initial structures were energy minimized with weak restraints on the heavy atoms to relieve any strain in the structure. The peptides were solvated in 3500 TIP3P [110, 112] water molecules. The CHARMM [114] forcefield with correction terms for dihedral angles [113] was used for the peptides.

For calculating h_α^{ex} (and thus s_α^{ex}) we adapted the hydration-shell-wise procedure

as noted in section 2.3 of page 21. Apart from this hydration shell method, we also calculated entropy of hydration directly from

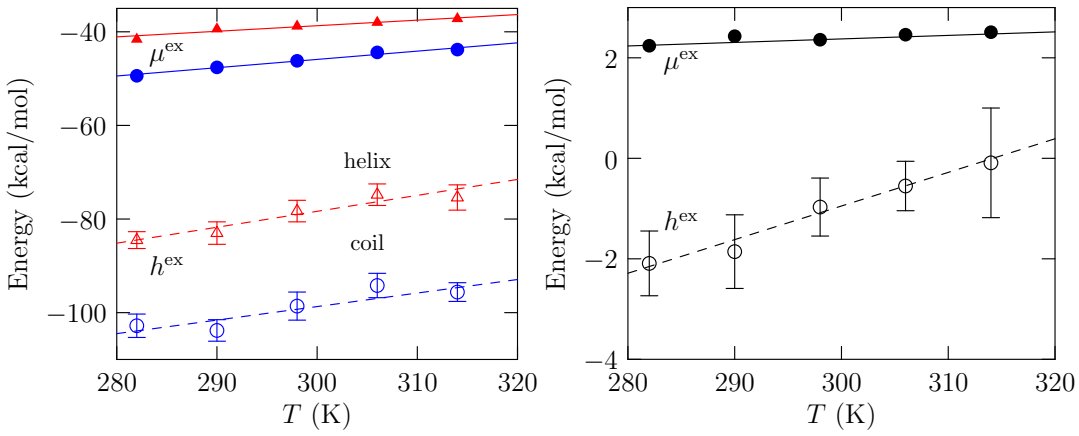
$$s^{\mathrm{ex}} = -\Big(\frac{\partial \mu^{\mathrm{ex}}}{\partial T}\Big)_{T,P} . \tag{3.1}$$

For this, we performed simulations at 280 K, 290 K, 306 K, and 314 K in addition to 298 K. We used a straight line fit to calculate entropy from this. The two different estimates of s^{ex} helped provide a check on the calculation of this rather difficult to assess quantity.

3.4 Results

3.4.1 μ^{ex} and h^{ex} *versus* T

Figure 3.1 (left panel) shows the temperature dependence of μ^{ex} and h^{ex} for the helix and coil states, while Fig. 3.1 (right panel) shows the same for methane. Table 3.1 collects the estimates of Ts^{ex} obtained from Eqs. 3.1 and 2.4. It is readily seen that within statistical uncertainties the agreement between these estimates is very good. The uncertainty in using Eq. 2.4 is high primarily because of the uncertainty in estimating h_{α}^{ex}; the uncertainty in using Eq. 3.1 is high because the small uncertainty in the slope (Fig. 3.1) is amplified by the factor T.



(a) Left Panel (b) Right Panel

Figure 3.1: Left panel: μ^{ex} (filled symbols) and h^{ex} (open symbols) for helix (triangle) and coil (circle) states. The standard error of the mean at 1σ is shown for h^{ex}. The standard errors for μ^{ex} are smaller than the size of the symbols. Right panel: corresponding data for methane.

Table 3.1: Estimate of Ts^{ex} ($T = 298.15$ K) obtained by fitting a straight line to μ^{ex}-*vs*-T (Eq. 3.1) and from Eq. 2.4 and estimate of the excess heat capacity of hydration c_p^{ex}. Ts^{ex} is in units of kcal/mol and c_p^{ex} is given in cal/mol/K. Standard errors of mean are given at 1σ.

	Ts_α^{ex} (Eq. 3.1)	Ts_α^{ex} (Eq. 2.4)	c_p^{ex}
Helix	-36.0 ± 6.0	-40.0 ± 4.0	341 ± 57
Coil	-53.0 ± 5.0	-52.0 ± 3.0	289 ± 82
Methane	-2.1 ± 0.6	-3.3 ± 0.6	67 ± 7

For hydrated-methane at 298 K, the calculated entropy of -7 cal/mol/K from temperature derivative of chemical potential or -11 cal/mol/K from Eq. 2.4 can be compared with the experimental value of -16 cal/mol/K [108, 109]. In an earlier study, with the SPC/E water model [115] and somewhat different parameters of

methane, it was found that $s^{\text{ex}} \approx -15$ cal/mol/K [71]. The poor numerical agreement with the current simulation is likely due to limitations in both the current water model and in the parameters of the solute. But, importantly, the sign of s_α^{ex} is correctly captured and the estimated Ts_α^{ex} is in error only by a few kcal/mol.

The heat capacity of hydration for the proteins and for methane is positive (Table 3.1). The experimental partial molar heat capacity of hydration at 298 K for methane is about 55 cal/mol/K [108], somewhat lower than the simulated value. On the other hand, experimentally determined enthalpy of hydration is about -2.7 kcal/mol versus the -1.0 kcal/mol obtained in simulations. These deviations reiterate the limitations of the water model and solute parameters.

We find that in the helix-to-coil transition the change in heat capacity is negative, about -52 cal/mol/K. The negative sign is opposite to what is normally observed in the unfolding of proteins [95, 102], but appears to be consistent with recent experimental data [117] and simulations results of helix-coil transition interpreted within the Zimm-Bragg [118] or Lifson-Roig [119] formalisms.

We conclude this section by emphasizing that the hydration of methane, a prototypical hydrophobe, and the peptides displays the two key signatures of hydrophobic hydration [120]: $s_\alpha^{\text{ex}} < 0$ and $c_p^{\text{ex}} > 0$. The sign of the excess entropy is usually rationalized by invoking some sort of water structuring around the solute. The sign of the heat capacity is then rationalized by suggesting that these ordered waters can absorb heat to melt and become disordered. This model has guided much of the thinking about hydrophobic hydration and its role in protein folding (see section 1.1.1 on page 4).

As we show below, a pictorial rationalization of the excess entropy and heat capacity of hydration in terms of ordering of water ignores the important role solute-solvent attractive interactions, the hydrophilic effects, play in the observed temperature dependence. To this end, we first consider the temperature dependence of subcomponents of h_α^{ex} and then, on the basis of the quasichemical decomposition, the

temperature dependence of the subcomponents of Ts_α^{ex}.

3.4.2 Temperature dependence of h^{ex} components

The average solute-water binding energy, $\langle E_\alpha \rangle$, and the solvent reorganization, $\langle E_{reorg} \rangle$, are approximately linear in T. For the proteins, we further decompose $\langle E_\alpha \rangle$ into contributions from the backbone and side chain, which also depend linearly on T. Thus the appropriate c_p^{ex} contribution (Table 3.2) is readily calculated.

Table 3.2: Contributions to c_p^{ex} from the temperature dependence of $\langle E_{reorg} \rangle$ and the subcomponents of $\langle E_{sw} \rangle$. For methane, the solute-water contribution is listed under side-chain water. For the peptide, E_α is partitioned into backbone-water and sidechain-water contributions. The net c_p^{ex} does not sum exactly to the values noted in Table 3.1 because the linear fits are not exact. All values are listed in cal/mol-K.

	Methane	Helix	Coil
Reorganization	60.0 ± 6.0	102.0 ± 51.0	-6.0 ± 79.0
Side-chain water	6.7 ± 0.7	47.0 ± 3.0	49.0 ± 2.0
Backbone water		188.0 ± 5.0	241.0 ± 4.0

For methane, the excess heat capacity of hydration is positive. Close to 90% of the value is accounted for by the temperature dependence of the reorganization term. As found before [71], the reorganization energy is positive. For an ordered first hydration shell of water molecules, one would expect a negative reorganization energy. The simulations suggest an absence of such a structure.

The heat capacity of hydration is positive for the helix and the coil states as well, but different physics accounts for the observed magnitude. For the helix about 70% of the heat capacity is accounted for by the temperature dependence of $\langle E_\alpha \rangle$, the hydrophilic contribution to hydration, and this proportion rises to essentially 100% for the coil.

3.4.3 Temperature dependence of $s^{\rm ex}$ components

In table 3.3 we present the chemistry, packing, and long-range contribution to the excess entropy. For the $R = 5$ Å case, the $s_\alpha^{\rm ex}$ components were obtained from the temperature derivative of the respective $\mu_\alpha^{\rm ex}$ components. To assess the sensitivity of the results, we also sought $s_\alpha^{\rm ex}$ components for $R = 3$ Å. The $R = 3$ Å inner-shell envelope hugs the molecular surface making the chemistry contribution to $\mu_\alpha^{\rm ex}$ zero. The packing contribution is easily calculated as before, but it is no longer possible to calculate the outer-term contribution using the Gaussian approximation. Since $s_\alpha^{\rm ex}$ is independent of R, we infer $s_\alpha^{\rm ex}$[outer] ($R = 3$ Å) from the calculated $s_\alpha^{\rm ex}$[pack]($R = 3$ Å) and the value of $s_\alpha^{\rm ex}$ obtained from the $R = 5$ Å data.

Table 3.3: Contributions to $Ts_\alpha^{\rm ex}$ at $T = 298.15$ K from Eq. 3.1. The sum of the individual $s_\alpha^{\rm ex}$ contributions does not add to precisely the same value as that from the net $\mu_\alpha^{\rm ex}$ as a linear fit with respect to T of the subcomponents of $\mu_\alpha^{\rm ex}$ is only an approximation. Since $s_\alpha^{\rm ex}$ is a state function, its value at $R = 5$ Å serves a reference in estimating the value of $s_\alpha^{\rm ex}$[outer] for $R = 3$ Å (indictated by $\star$). All values are listed in kcal/mol.

	$R = 5$ Å			$R = 3$ Å		
	Helix	Coil	Methane	Helix	Coil	Methane
$Ts_\alpha^{\rm ex}$[pack]	32 ± 4	43 ± 5	0.4 ± 0.7	13 ± 4	7 ± 3	-0.6 ± 0.1
$Ts_\alpha^{\rm ex}$[chem]	-50 ± 3	-74 ± 2	-1.4 ± 0.6	0	0	0
$Ts_\alpha^{\rm ex}$[outer]	-18 ± 4	-23 ± 1	-1.2 ± 0.1	$-48^\star$	$-60^\star$	$-1.5^\star$
$Ts_\alpha^{\rm ex}$ (LHS Eq. 3.1)	-35 ± 6	-53 ± 4	-2.1 ± 0.7	-35	-53	-2.1

Consider first the case of methane. For $R = 3$ Å, the hydrophobic contribution, $s^{\rm ex}$[pack], is negative, a result that has traditionally been interpreted in terms of ordering of water molecules around the cavity, as a way for the water to satisfy its hydrogen bonding. But the negative -1.5 kcal/mol arising from the interaction of the solute with the solvent is decidedly due to solute-solvent attractive interactions.

These difficulties become acute when we consider $Ts_\alpha^{\rm ex}[\text{pack}]+Ts_\alpha^{\rm ex}[\text{outer}]$ for $R = 5$ Å case. The solute in this case is special in excluding water molecules up to a range R, and $s^{\rm ex}$ for this solute is itself negative, challenging any structural rationalizations.

The comparison of methane with the helix and coil sharpens these observations. For a *single* $R = 3$ Å cavity, $s_\alpha^{\rm ex}[\text{pack}] < 0$, but for a union of such cavities that conform to the helix or coil shapes, $s_\alpha^{\rm ex}[\text{pack}] > 0$. We find that the net negative sign of $s_\alpha^{\rm ex}$ arises from the hydrophilic contributions to $\mu_\alpha^{\rm ex}$. In effect, with increasing temperature, the effective solute-solvent attraction is made less favorable, elevating $\mu_\alpha^{\rm ex}$, $h_\alpha^{\rm ex}$, and $-s_\alpha^{\rm ex}$.

3.5 Summary

Our study leads to several important conclusions. First, the hydration thermodynamics of a small cavity (an ideal hydrophobe) is entirely irrelevant in inferring the hydration thermodynamics of a macroscopic cavity; to wit, even the sign of the excess entropy of hydration differs. Thus translating notions of hydrophobicity, be it in experimental or modeling approaches, from small molecules (or small length scale) to infer the hydrophobic hydration of a macromolecule is inappropriate. But this is precisely the approach — particularly the peculiar temperature signatures in protein unfolding and protein-protein association and its similarity in small molecular compounds — that led to early suggestions on hydrophobicity as a potentially important driving force in protein folding. Subsequent researches elevated hydrophobicity to the status of a dominant force in protein folding, but a common theme in these studies is the reliance on translating notions of hydrophobicity from the small length scale to a protein.

Often entropy and disorder are used interchangeably. For thermodynamics of interest in biophysics, such notions have merit when the system energy is a constant; then one can talk of entropy in terms of arrangements of the constituent parts of

the system. But when the system is undergoing volume changes (to keep pressure constant) and energy exchange (to keep temperature constant), interpreting entropy in terms of arrangements is not possible. This is, of course, just a standard result of statistical mechanics, but one that is surprisingly overlooked in pictorial rationalizations of the excess entropy of hydration. We show that both $s_\alpha^{\rm ex} < 0$ and $c_p^{\rm ex} > 0$ have a very simple explanation that does not require us to invoke water structuring around hydrophobic groups. With increasing temperature the effective attractive interaction of the solute with the solvent is weakened, elevating the excess chemical potential of the solute: thus $s_\alpha^{\rm ex} < 0$. Likewise, $c_p^{\rm ex}$ is positive because the excess enthalpy of hydration also increases (becomes less favorable) with weakening of the effective solute-solvent attraction. The sign of both these quantities is directly related to hydrophilic effects and not due to hydrophobic effects.

We identify three factors that contribute to $c_p^{\rm ex} > 0$: backbone-solvent interactions, side-chain solvent interactions, and water reorganization. The contribution of the backbone-solvent and side-chain solvent interaction to $h_\alpha^{\rm ex}$ is negative, but increases with temperature. The reorganization factor is positive and increasing with temperature. For the peptides, changes in the the backbone-solvent contribution are by far the most dominant contribution to $c_p^{\rm ex}$. Interestingly, the importance of backbone hydration suggested by our study is in consonance with the identification of the importance of backbone hydration in explaining protein stability changes due to added osmolytes. While the studies identifying the importance of backbone hydration also rely on an additive model of free energy change in water to osmolyte solution transfer, we have recently suggested a plausible explanation why the additive model might work in that context (see chapter 5).

Lastly, our identification of hydrophilic effects in explaining $s_\alpha^{\rm ex} < 0$ and $c_p^{\rm ex} > 0$ also suggests a plausible explanation for cold denaturation. With decreasing temperature, the hydration of the unfolded state is likely preferred over the folded state because of hydration of the peptide backbone, which is found to be important based

on the c_p^{ex} data.

4

Chemical stresses in protein solutions modulate attractive interactions

Note: This chapter draws upon a manuscript in preparation.

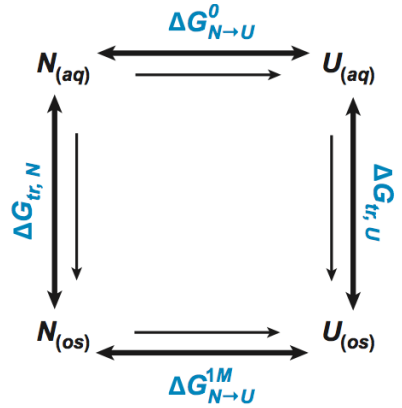
4.1 Overview

In efforts to understand the thermodynamic forces stabilizing a protein, recourse has often been taken to modulate the solvent using denaturants. Interestingly, in studies on adaptation of organisms to their environment, it was discovered that, in the face of mechanical, thermal, and chemical stresses, all organisms actively tune the solvent by synthesizing osmolytes to preserve protein structure. Studies within the last 15 years have led to a thesis that osmolytes primarily act by changing the solubility of the peptide backbone [43, 44]. In chapter 5, we will study the hydration of a peptide unit in osmolytes, but here we illustrate the physical clarity of our framework by studying the helix-coil transition of a model peptide in several osmolytes. A broader consequence of our work is that attractive interactions are more important

in protein solvation thermodynamics than the hydrophobic effect.

4.2 Introduction

In the Tanford transfer model, transfer free energies of peptides (transfer free energy being free energy of transferring a peptide from water to aqueous solution of osmolytes) are needed to interpret m-values (an indicator of relative stability of protein conformations) of protein folding (Figure 4.1).



$$m = \Delta G^{1M}_{N\to U} - \Delta G^{0}_{N\to U} = \textstyle\sum_i \alpha_i \Delta g_{\mathrm{tr},i}$$

$$\Delta g_{\mathrm{tr,\,i}} \Rightarrow \text{group transfer free energy}$$

Figure 4.1: Tanford transfer model (figure taken and modified from reference [93]). α_i are surface area based factors intended to translate group transfer free energies, ($\Delta g_{tr,i}$), from model compounds to proteins.

The m-value interpreted within the group-additive decomposition of the transfer free energy leads to the observation that denaturants and osmolytes tune protein stability primarily by their effect on the protein backbone [43, 44]. Interpreting these transfer free energies in terms of preferential solvation framework (protecting osmolytes must partition away from the protein while denaturing osmolytes partition

closer to the protein [81]) have also led to the "osmophobic effect" as yet another thermodynamic force governing protein stability [94]!

Interestingly, a recent reanalysis of Auton and Bolens group-transfer free energies by Moeser and Horinek [97] suggests that the group-transfer contributions calculated by Bolen and coworkers were itself in error in ignoring role of concentration units, a conclusion that while true likely does not change the transfer free energies relative to glycine (Bolen/Pettitt; personal communication). Their reanalysis leads Moeser and Horinek to conclude that the transfer of all groups (not just the peptide bond) from water to aqueous urea is favored.

The group-additive model of m has greatly advanced the discussion of efforts to understand the role of osmolytes in protein stability, but many aspects of the folding/unfolding of proteins in aqueous osmolytes remains puzzling. For example, urea is known to raise surface tension but it can stabilize states of higher surface area (i.e. unfolded states); trimethylamine N-oxide (TMAO) on the other hand lowers surface tension but it stabilizes states of low surface area (i.e. the folded state). Other experiments further contribute to the puzzle: Li and Walker have found that surface tension plays a dominant role in protein folding in presence of osmolytes [87] but their later experiments with temperature dependence showed signatures of hydrophobicity [88]; here it should be noted that hydrophobicity and surface tension have different temperature dependence.

Given the importance of protein denaturation, many computer simulation studies have sought to explore the role of urea in protein unfolding. For example, simulation study by Daggett and coworkers [89] suggests that the protein first swells in aqueous urea and unfolds after subsequent penetration by urea into the protein. Thirumalai and coworkers have explored the role of urea and over papers spanning a decade have suggested that urea unfolds because of its role in electrostatics [90], in decreasing the hydrophobic effect [91], and more recently by binding to the protein [92]. (We note here that that given the role of urea in increasing the surface tension, one must expect

urea to increase the magnitude of primitive hydrophobic effects.) Study from Pettitt's group [98] supports the thesis that urea denaturation is more due to dispersion interactions, but these simulations definitely suffered from convergence problems [99].

While a consensus appears to be developing for the role of urea (at least within simulations), some of the puzzles noted above remain unresolved. Further, given the obvious limitations of the group-additive model (chapter 5), a reexamination based on our tools might prove helpful.

In an effort to understand the thermodynamics of protein solvation to resolve some of these puzzles, in this work we will look how some cosolutes change solubility of a few smaller peptides. We will later discuss its implications for protein folding. We have studied effect of urea, TMAO and guanidinium hydrochloride on solubility of smaller peptides using the framework developed in this thesis. Urea and guanidinium chloride can denature a protein [75] while TMAO is known to refold a denatured protein.

We will next discuss our simulation methods followed by results and discussion.

4.3 Methods

We take deca-glycine and deca-alanine in helical and coiled conformations. For these peptides, we have calculated hydration free energies and solvation free energies in aqueous solutions of TMAO, urea, and guanidinium hydrochloride. We will use these to calculate transfer free energies of peptides.

While transfer free energies can easily give m-value of protein folding, calculating the transfer free energy of the unfolded state requires thermal averaging over entire ensemble of unfolded state; here we will take an extreme model of the unfolded state (a fully exposed conformation). We also note a slight difference of notation that we have followed in our thesis in the definition of m-value. m-values are defined for a 1 molar solution of the cosolute, but in our thesis we have not used 1 molar solutions.

For a solution, the m-values presented in this thesis are not based on one molar solution basis but on the basis of actual concentration of that solution.

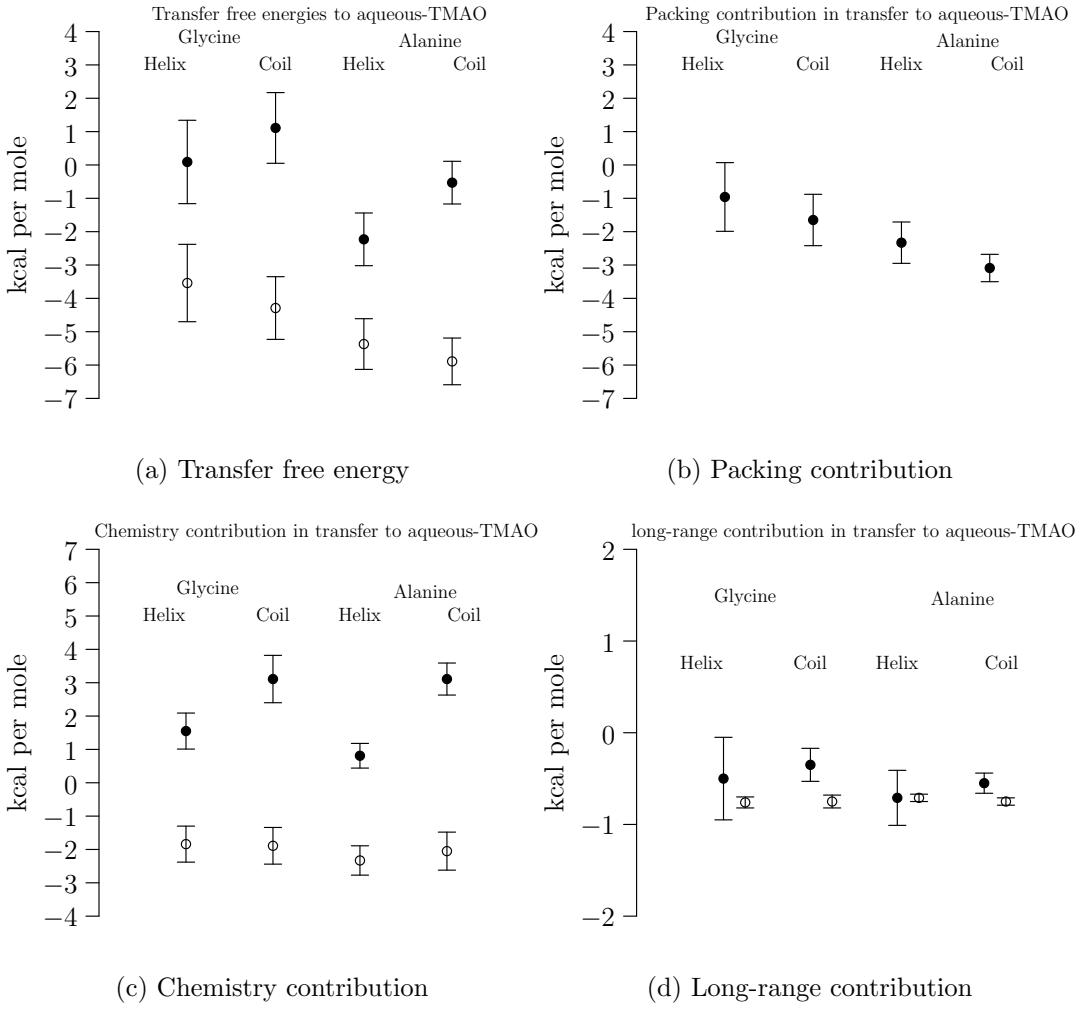
The TIP3P model of water [110, 112] is used and deca-peptides are modeled by a CHARMM force field [114]. 340 TMAO molecules modeled by Kast force-field [79] are present with 3500 water molecules to give an approximately 4 molar solution. The approximate size of the box in this system if 51.86 Å. 783 urea molecules are mixed with 3500 water molecules to make an approximately 8 molar solution of urea. The size of the box in this case it approximately 54.36 Å. 500 molecules of guanidinium hydrochloride are mixed with 3500 water molecules in a box of approximate size 54.48 Å to give an approximately 5 molar solution of GDCL. Urea and GDCL are modeled by Kirkwood-Buff based fore-field of Weerasinghe and Smith [77, 78]. Rest of the methods are as listed in section 2.3 on page 21.

4.4 Results

4.4.1 helix-coil transition in aqueous-TMAO solution

Figure 4.2 presents transfer free energies to the aqueous solution of TMAO.

The sign of transfer free energy is not uniform for all the peptides (osmophobicity requires that transfer free energies be positive for all peptides to a solution of osmophobe like TMAO [94]). However, all discharged peptides have negative transfer free energies hinting that the charge on the peptides should cause the non-uniformity of peptide transfer free energy. Moreover, packing contributes to make the transfer free energies negative (consistent with experimental observation on surface tension, decrease in packing here is found to be proportional to surface area), long-range effects also contribute to make the transfer free energy negative. But, the short-range



(a) Transfer free energy

(b) Packing contribution

(c) Chemistry contribution

(d) Long-range contribution

Figure 4.2: Transfer of peptides from pure water to aqueous-TMAO solution. Filled symbols are for fully charged peptides, empty symbols are for discharged peptides

attractive interactions (chemistry) is always positive for a fully charged peptide and always negative for a discharged peptide. In fact, higher the exposure of the backbone to the solvent, more positive the transfer free energy. Interestingly if the peptide is discharged, TMAO will fail to stabilize the helical state of peptide. Thus TMAO must act by "dialing down" [93] the short-range attractive interactions with the solvent.

In Table 4.1 we present the m-value of of the charged and discharged peptides in TMAO. It is evident that if the peptide-solvent electrostatic interactions are not

Table 4.1: m-value and its components for aqueous solution of TMAO. Errors are shown at the 2σ level.

Peptide	Discharged	Packing	Chemistry	Long-range	μ^{ex}
Glycine	No	-0.69 (1.29)	1.56 (0.89)	0.15 (0.48)	1.02 (1.64)
	Yes	-0.69 (1.29)	-0.05 (0.77)	0.01 (0.09)	-0.73 (1.50)
Alanine	No	-0.76 (0.74)	2.30 (0.61)	0.16 (0.32)	1.70 (1.02)
	Yes	-0.76 (0.74)	-0.28 (0.72)	0.04 (0.06)	-1.00 (1.03)

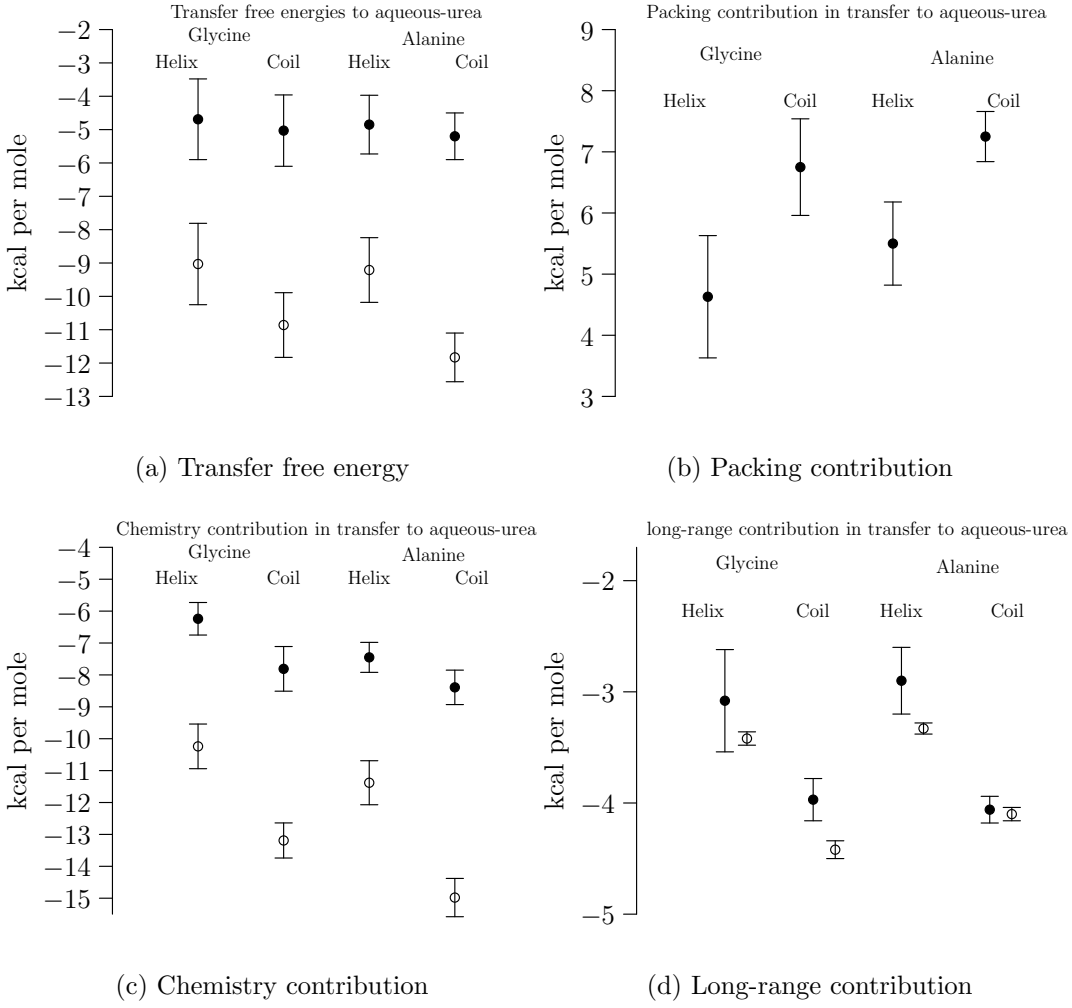
allowed by discharging the peptide, then TMAO can not refold it. In fact, for a discharged peptide, TMAO has a tendency to unfold. And we have seen in Figure 4.2 and Table 4.1 that TMAO "weakens" local attractive interactions. Our suggestion is that TMAO acts by "weakening" the short-ranged electrostatic interactions between the the peptide and the solvent; backbones-solvent interactions are likely to contribute significantly as a side-chain has zero electrostatic interaction with the solvent.

4.4.2 helix-coil transition in aqueous-urea solution

Figure 4.3 collects the free energies of transferring deca-glycine and deca-alanine from water to urea. Unlike aqueous-TMAO, aqueous-urea has consistent effect on transfer free energy: it stabilizes the peptides, and it stabilizes coiled states even more.

Moreover, urea increases hydrophobic effect. But, urea also enhances short-range attractive interactions and long-range van der Waals .

Using these transfer free energies, we calculate the m-value in aqueous-urea. These values are listed in Table 4.2. As is expected of an osmolyte that increases hydrophobicity, the contribution of the hydrophobic effect to the m-value is toward folding the peptides. But, attractive interactions also contribute significantly. However, quite surprisingly the effect of long-range non-specific van der waals in the m-value is not insignificant – i.e. urea has a preference for a conformation even when it is not allowed



(a) Transfer free energy

(b) Packing contribution

(c) Chemistry contribution

(d) Long-range contribution

Figure 4.3: Transfer of peptides from pure water to aqueous-urea solution. Filled symbols are for fully charged peptides, empty symbols are for discharged peptides

to hydrogen bond.

Urea is a better denaturing agent for discharged peptides (Table 4.2). In fact, when charges are turned on, urea becomes poorer denaturing agent for the same peptide. Thus, urea denatures a protein purely by dispersion interactions, these dispersion interactions come from both the backbone and side-chain.

Rather interestingly, if alkanes are to be taken as models for hydrophobic effect, then urea which strengthens the hydrophobic effect must protect the helical state for

Table 4.2: m-value and its components for aqueous solution of urea. Errors are shown at the 2σ level.

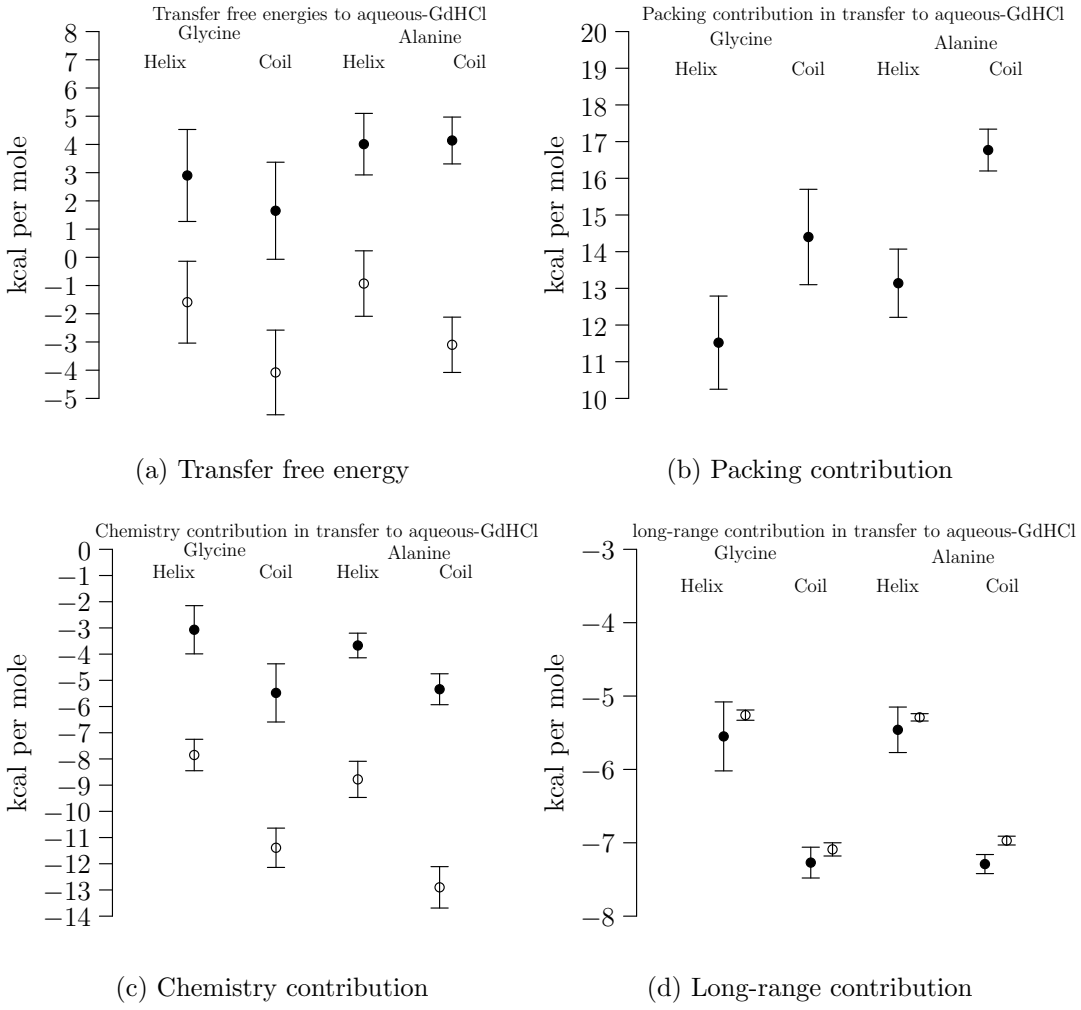
Peptide	Discharged	Packing	Chemistry	Long-range	μ^{ex}
Glycine	No	2.12 (1.27)	-1.57 (0.87)	-0.89 (0.50)	-0.34 (1.62)
	Yes	2.12 (1.27)	-2.95 (0.90)	-1.00 (0.10)	-1.83 (1.56)
Alanine	No	1.75 (0.80)	-0.94 (0.72)	-1.16 (0.32)	-0.35 (1.12)
	Yes	1.75 (0.80)	-3.60 (0.91)	-0.77 (0.08)	-2.62 (1.21)

discharged peptides. But in our study we observe that urea stabilizes coiled state of discharged peptides. Thus, alkanes are not good models to study hydrophobicity, because they too have attractive dispersion interactions with the solvent.

4.4.3 helix-coil transition in aqueous-GdHCl solution

We present transfer free energies in Figure 4.4. Interestingly for an osmolyte that acts as a denaturing agent, transfer free energies are positive. Guanidinium hydrochloride (GdHCl) raises chemical potentials by increasing packing contribution. Experiments have shown that GdHCl increases surface tension. Besides, increasing surface tension, GdHCl also has attractive interactions with the peptide. These attractive interactions are higher for a discharged peptide. Additionally, long-range attractive interactions with the peptide are better in GdHCl as compared to pure water.

Comparing the transfer free energies of helical and coiled peptides, we learn GdHCl is a better unfolding agent for discharged peptides. In fact, chemistry of a discharged peptide is more negative than the chemistry of a fully charged peptide. We suggest that GdHCl unfolds proteins by dispersion interactions. This suggestion has been made before us [80].



(a) Transfer free energy

(b) Packing contribution

(c) Chemistry contribution

(d) Long-range contribution

Figure 4.4: Transfer of peptides from pure water to aqueous-GdHCl solution. Filled symbols are for fully charged peptides, empty symbols are for discharged peptides

4.5 Summary

The role of long range dispersion interaction is often neglected in transfer free energy and m-value calculations [93]. But we have seen here, that long-range dispersion interactions play significant role in transfer free energies and m-value. However, long-range electrostatic interactions are not modulated by any of the osmolytes studied in this work.

Table 4.3: m-value and its components for aqueous solution of GdHCl. Errors are shown at the 2σ level.

Peptide	Discharged	Packing	Chemistry	Long-range	μ^{ex}
Glycine	No	2.88 (1.81)	-2.41 (1.44)	-1.72 (0.51)	-1.25 (2.38)
	Yes	2.88 (1.81)	-3.54 (0.95)	-2.41 (0.11)	-2.49 (2.08)
Alanine	No	3.63 (1.09)	-1.67 (0.75)	-1.83 (0.34)	0.13 (1.37)
	Yes	3.63 (1.09)	-4.12 (1.05)	-1.68 (0.08)	-2.17 (1.51)

Besides the neglect, the range at which various attractive interactions are active is also misunderstood in the literature. It is assumed that dispersion interaction are short ranged and osmolytes should come close to the peptide in order to have better dispersion interactions [98]. Similarly, it is also assumed that because electrostatic interactions are long-range, a change in long-range attractive interactions will mean a change in electrostatic interactions. Our results differ. Dispersion interactions contribute significantly to long-range attractive interactions and they can be modulated by osmolytes. We find that osmolytes can change dispersion interactions even when they are beyond the first hydration shell of the solute – dispersion interactions are short ranged between two molecules, they are not short ranged between a molecule and a big bath of molecules. (A quick estimate for interaction between a molecule and a big bath of molecules: the integration of $\frac{1}{r^6}$ term of dispersion interaction is performed after multiplying it with a $4\pi\ r^2 dr$ term; the latter term is monotonically increasing in r.)

The backbone based theory [43, 44] suggests that osmolytes primarily act on the backbone, osmolytes that fold a protein raise chemical potential of both the native and unfolded state, and osmolyte that unfold lower the chemical potential of both native and unfolded state. This theory is not entirely supported by our work. We observe in the case of TMAO, though it acts via backbone based mechanism, all conformations did not see a rise in chemical potential. We observe in the case of urea that unfolding

is a consequence of promiscuous interactions by urea. And we observe in the case of guanidinium hydrochloride that osmolytes that denature a protein can also raise chemical potential of protein's conformations. Broadly, our work finds that attractive interactions are more important in protein solution thermodynamics than the hydrophobic effect – chapter 3 also highlights importance of attractive interactions.

5

Limitations of group additivity

Note: This chapter draws upon the following two published papers: (1) D. S. Tomar, D. Asthagiri, V. Weber. "*Solvation Free Energy of the Peptide Group: Its Model Dependence and Implications for the Additive-Transfer Free-Energy Model of Protein Stability*". Biophysical Journal, Volume 105, Issue 6, Page 1482; (2) D. S. Tomar, V. Weber, B. M. Pettit, and D. Asthagiri. "*Conditional Solvation Thermodynamics of Isoleucine in Model Peptides and the Limitations of the Group-Transfer Model*". Journal of Physical Chemistry B. Volume 118, Issue 15, Page 4080.

5.1 Overview

In Chapters 3 and 4, based on the framework we developed, we showed distinct limitations and/or flaws in the canonical explanations of the physics of protein hydration. Since the earliest studies on proteins, including techniques implicit in several modern approaches to model protein hydration [19, 43], the thermodynamics of protein hydration (or solvation) has been obtained by using a group-additive scheme (Figure 5.1).

Here we explore this approximation first for a hydrophilic group (the backbone) and then hydrophobic group (butane-*analog*). We uncover subtle, but consequential,

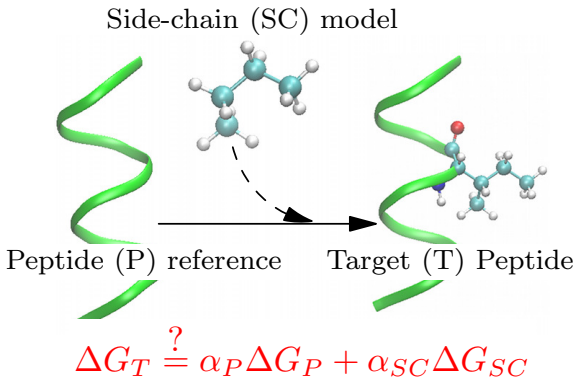


Figure 5.1: Pictorial representation of the group additivity scheme, here shown for a side-chain. This scheme assumes that the transfer free energy of the target peptide (ΔG_T) is a weighted sum of transfer free energies of the side-chain (ΔG_{SC}) and the reference peptide (ΔG_P).

flaws in this group-additive approach. Our work thus suggests much caution needs to be exercised in using such group-additive approaches to model protein solution thermodynamics and the mechanism of protein folding.

5.2 Introduction

The group-additive decomposition of the unfolding free energy of a protein in an osmolyte solution relative to that in water poses a fundamental paradox: whereas the decomposition describes the experimental results rather well, theory suggests that a group-additive decomposition of free energies is, in general, not valid. In a step towards resolving this paradox, here we study the peptide-group transfer free energy. We calculate the vacuum to solvent (solvation) free energies of $(\mathrm{Gly})_n$ and cyclic diglycine (cGG) and analyze the data following experimental protocol. The solvation free energies of $(\mathrm{Gly})_n$ are linear in n suggesting group-additivity. However, the slope interpreted as the free energy of a peptide unit differs from that for cGG scaled by a factor of half, emphasizing context dependence of solvation. But the water-

to-osmolyte transfer free energies of the peptide unit are relatively independent of the peptide model, as observed experimentally. To understand these observations, a way to assess the contribution to the solvation free energy of solvent-mediated correlation between distinct groups is developed. We show that linearity of solvation free energy with n is a consequence of uniformity of the correlation contributions, with apparent group-additive behavior in the water-to-osmolyte transfer arising due to their cancellation. Implications for inferring molecular mechanisms of solvent-effects on protein stability on the basis of the group-additive transfer model are suggested.

Moreover, the hydration thermodynamics of the amino acid X relative to the reference G (glycine) or the hydration thermodynamics of a small molecule analog of the sidechain of X is often used to model the contribution of X to protein stability and solution thermodynamics. We consider the reasons for successes and limitations of this approach by calculating and comparing the conditional excess free energy, enthalpy, and entropy of hydration of the isoleucine sidechain in zwitterionic isoleucine, in extended penta-peptides, and in helical deca-peptides. Butane in *gauche* conformation serves as a small molecule analog for the isoleucine sidechain. Parsing the hydrophobic and hydrophilic contributions to hydration for the sidechain shows that both these aspects of hydration are context-sensitive. Further, analyzing the solute-solvent interaction contribution to the conditional excess enthalpy of the sidechain shows that what is nominally considered a property of the sidechain includes entirely non-obvious contributions of the background. The context-sensitivity of hydrophobic and hydrophilic hydration and the conflation of background contributions with energetics attributed to the sidechain limit the ability of a single scaling factor, such as the fractional solvent exposure of the group in the protein, to map the component energetic contributions of the model-compound data to their value in the protein. But ignoring the origin of cancellations in the underlying components the group-transfer model may appear to provide a reasonable estimate of the free energy for a given error tolerance.

5.3 Protein backbone

The thermodynamics of protein unfolding in the presence of aqueous osmolytes (small organic co-solutes) is of fundamental interest in understanding the forces stabilizing the folded protein and in the broader quest to understand how biological systems adapt to environmental stresses by changing the solvent properties.

Experimental investigations on the molecular role of the osmolytes on the thermodynamics of protein unfolding are usually anchored by

$$\Delta G_{NU} = \Delta G^{(0)}_{NU} + \sum_i \alpha_i \Delta g_{\mathrm{tr},i}\,, \tag{5.1}$$

where ΔG_{NU} and $\Delta G^{(0)}_{NU}$ are, respectively, the experimentally accessible unfolding free energies of the protein in the osmolyte solution and in water. (For a 1 M osmolyte solution, $\Delta G_{NU} - \Delta G^{(0)}_{NU}$ is the m-value; here we follow the sign-convention of Ref. 93.)

Drawing upon the studies by Tanford and coworkers, $\Delta G_{NU} - \Delta G^{(0)}_{NU}$ is partitioned into water-to-osmolyte solution transfer free energy $\Delta g_{\mathrm{tr},i}$ contribution due to group i, potentially allowing one to understand the role of the solvent at the level of each individual group. Typically, the groups include the peptide unit and the side-chains of the amino acid residues, and the $\Delta g_{\mathrm{tr},i}$ contributions are obtained from chemically apposite model compounds. The factor α_i accounts for the fractional change in the degree of solvent-exposure of group i in going from the native (N) to unfolded (U) state.

The pioneering experimental studies by Bolen and coworkers interpreted using Eq. 5.1 leads to an important insight: both conformation-protecting and denaturing osmolytes exert their influence primarily by changing the solubility of the peptide backbone [93, 94], an identification with significant consequences to our understanding of protein folding [43, 44]. However, these results pose a paradox: whereas the group-additive decomposition in Eq. 5.1 is, in general, not valid, $\Delta G_{NU} - \Delta G^{(0)}_{NU}$ calculated using Eq. 5.1 together with reasonable approximations of the unfolded state of the

protein describes the experimental $\Delta G_{NU} - \Delta G_{NU}^{(0)}$ rather well. Towards resolving this paradox, here we examine the vacuum-to-solvent and water-to-osmolyte solution transfer free energies of the peptide group and its dependence on the choice of model compounds.

Glycyl-peptides, $(\mathrm{Gly})_n$, are a common model for the peptide unit. Using blocked glycyl peptides and by carefully minimizing peptide-peptide interactions in peptide solubility, Auton and Bolen [74] have sought the peptide group transfer free energy, $\Delta g_{\mathrm{tr},p}$, that is "independent of the model compound and the choice of concentration scale". In their studies, $\Delta g_{\mathrm{tr},p}$ was obtained by appropriately combining the transfer free energy of chains of various lengths n, and, somewhat more robustly, by equating $\Delta g_{\mathrm{tr},p}$ to the slope of the transfer free energy with respect to n, the so-called constant increment method. Model independence was demonstrated by showing that the transfer free energies based on $(\mathrm{Gly})_n$ agree with those from cGG, the cyclic-diglycine molecule. This concordance, while pleasing, is puzzling as well, for in cGG the CO and NH of the peptide are *cis* and the molecule has a net zero dipole moment, whereas the CO and NH are *trans*, and the peptide dipole moment non-negligible, in the $(\mathrm{Gly})_n$ system. Thus, either the conformation of the peptide is unimportant in Δg_{tr} (Eq. 5.1) or there are other factors that lead to this result or a combination of both.

To test the group-additive transfer free energy model for peptides, a recent all-atom simulation study examined the free energy of $(\mathrm{Gly})_n$ ($n = 2 \ldots 5$). In this study, both the vacuum-to-water and the water-to-osmolyte solution transfer free energies were reasonably linear with n, an observation that was used to conclude in favor of group additivity. (No attempt was made to study cGG and thus also verify model independence.) However, using a continuum dielectric model of the solvent, Avbelj and Baldwin [153, 154] have argued against group additivity. For example, they find that in an $(\mathrm{Ala})_9$ chain, the electrostatic contribution to the hydration free energy of peptide unit as well as its interaction with other peptide units depends on the

location of the chosen peptide unit. As we shall show below, the preceding seemingly contradictory conclusions from the earlier studies can be reconciled; and this reconciliation depends on appreciating that an independent group additive contribution to the hydration free energy is not a consequence even when the conditions for use of the constant increment approach are satisfied.

Here we use theory and computer simulations to examine the solvation free energies of acetyl-$(Gly)_n$-methyl amide peptides and of cGG. The free energies are obtained by a quasichemical organization of the potential distribution theorem as discussed in section 2.1 on page 16. A virtue of this formulation is that it makes transparent the role of correlated fluctuations of the binding energies of two groups on the molecule and its role in the thermodynamics of hydration. We show that even for an idealized solute that is incapable of making any close (near-neighbor) contacts with the solvent, the group-solvent binding energies between neighboring groups are correlated and make a non-negligible contribution to the net solvation free energy. This implies that identifying a model independent, group-contribution to the solvation free energy is, in principle, not possible, even for this idealized solute. The situation for a real solute is bound to be considerably more complicated.

Group additive behavior in the solvation free energy can be suggested if solvent-mediated correlation between different groups are similar for a series of model compounds, such as $(Gly)_n$, but the identified group-additive free energy will necessarily depend on the chosen model. Importantly, such correlation effects arise at different energy scales: within the quasichemical formulation it is found for the idealized solute, a collection of hard-spheres, and for the physical solute. However, group additivity in the water-to-osmolyte solution transfer free energy can arise if these correlation contributions cancel, but then care is needed in inferring mechanistic conclusions about solvent effects on stability from such an apparently group-additive model.

We will now review methods that we have employed in this study in addition to the methods listed in section 2.3 on page 21, after this we will present results of this

study.

5.3.1 Methods

The peptides are modeled in the extended configuration with the long axis aligned with the diagonal of the simulation cell and the center of the peptide placed at the center of the simulation cell. (Initial configurations were energy minimized with restraints to keep the peptide extended.) The peptide atoms are held fixed throughout the simulation. The solvent was modeled by the TIP3P [110, 112] model and the CHARMM [114] forcefield with correction terms for dihedral angles [113] was used for the peptide. A total of 2006 TIP3P molecules solvated the peptide. Parameters for urea and TMAO were obtained from Refs. 77 and 79, respectively. A total of 449 urea molecules (for a molar concentration of about 8 M) and 195 TMAO molecules (for a molar concentration of about 4 M) were used.

The long-range contribution $\mu_n^{\text{ex}}[P(\varepsilon|\phi_R)]$ was obtained by inserting the solute in a cavity (with atom-centered radius $R = 5$ Å). 1500 equally spaced cavity configurations were obtained from the last 0.375 ns of a 1 ns simulation at $R = 5$ Å. (The starting configuration for the $R = 5$ Å simulation was obtained from the endpoint of the Gauss-Legendre procedure as in section 2.3 on page 21.) The binding energies for the correlation analysis were obtained from the solute extraction procedure.

Cyclic-diglycine was built and optimized using the Gaussian (G09) quantum chemistry package [111]. For consistency with the $(\text{Gly})_n$ simulations, the partial charges and Lennard-Jones interaction parameters were obtained from the backbone atoms of the CHARMM forcefield.

5.3.2 Results

The excess free energies of solvation (as well as packing, chemistry, and long-range contributions) are linear in the peptide chain length in all the solvents studied here.

Table 5.1: Peptide group transfer free energies from vacuum to solvent obtained from the slope of $\mu_n^{\rm ex}$ versus n. Values for cGG have been scaled by 1/2. Below each line for the model system studied, we present the free energy values for transferring from water to the solution under study. All values are in kcal/mol. Standard error of the mean is about 0.1 kcal/mol (1σ).

	Water	Urea	TMAO
$(\mathrm{Gly})_n$	-5.0	-5.4	-5.0
		-0.4	0.0
cGG/2	-6.2	-6.6	-6.2
		-0.4	0.0

The water-to-aqueous osmolyte transfer free energy agrees quite well for both the $(\mathrm{Gly})_n$ and cGG models. The urea concentration is about 8 M and assuming a linear dependence of transfer free energy on osmolyte concentration [39, 75], we find that for 1 M urea solution, the transfer free energy is -50 ± 13 cal/mol, a value that is in good agreement with the experimental estimate of -43 ± 10 cal/mol [74]. We find a net zero transfer free energy to aqueous TMAO solution (4 M), in contrast to the experimental estimate of 87 cal/mol [74], a discrepancy that is likely due to inadequacy in the forcefield model for TMAO [39, 76].

From Table 5.1 we can note that the good agreement between $(\mathrm{Gly})_n$ and cGG in water to aqueous osmolyte transfer free energy masks the rather poor agreement in transfer free energies from vacuum to the respective solution. While it can be argued that water-to-osmolyte solution transfer is the most relevant experimentally, from the perspective of a physical theory, the vacuum-to-solution transfer quantities have the virtue of highlighting the role of inter-group correlations transparently.

To further illuminate the model dependence of the transfer free energies (Table 5.1), we consider how these values are used in modeling the m-value. As Eq. 5.1

indicates, the transfer free energies are scaled by the fractional solvent exposure of the group relative to that in the model compound [152]. Using commonly used atomic radii [155–157] for calculating solvent accessible surface area (SASA), we find that the SASA of the $CONHCH_2$ group in cGG is about 1.5 times that in the $(Gly)_n$ model. (The SASA of the peptide in cGG is obtained by dividing the SASA for the entire molecule by 2. For the $(Gly)_n$ model, the change in SASA with n gives the SASA per peptide.) Thus relative to $(Gly)_n$, the transfer free energy per unit area of the model compound is nearly 33% smaller in magnitude for the peptide in cGG. In the context of the m-value, using the peptide from cGG as a model can lead to both quantitative and qualitative errors.

Now, to understand the mechanistic aspects in a transfer free energy, we will first start by analyzing binding energies for a simple system, later we will study hydrophobic and short ranged attractive contributions.

5.3.2.1 Correlations in long range binding energies

Imagine a polymer that has n units, the binding energy (ε) of this polymer is summation of binding energies of all the units ($\varepsilon = \sum_i^n \varepsilon_i$, where ε_i is the binding energy of the unit $i \in (1, n)$.) Similarly, variance of the binding energy of that polymer can also be written in terms of the variance and covariances in the binding energy of the constituent units. ($\sum_i \langle \delta\varepsilon_i^2 | \phi_R \rangle + 2 \sum_{i>j} \rho_{ij} \sqrt{\langle \delta\varepsilon_i^2 | \phi_R \rangle \langle \delta\varepsilon_j^2 | \phi_R \rangle}$, where ρ_{ij} is the correlation coefficient [70].) Thus the long-range contribution of a polymer can be written in terms of units of that polymer.

$$\mu^{\text{ex}}[P(\varepsilon|\phi_R)] \;=\; \sum_i \mu^{\text{ex}}[P(\varepsilon_i|\phi_R)] + \beta \sum_{i>j} \rho_{ij} \sqrt{\langle \delta\varepsilon_i^2 | \phi_R \rangle \langle \delta\varepsilon_j^2 | \phi_R \rangle}\,, \qquad (5.2)$$

Necessary, but not sufficient, condition for context independence to hold is that the long-range contributions must be independent of the neighboring units; Equation, 5.2 makes it easy to check if this condition holds.

Table 5.2: Contributions to the outer term by a peptide unit for $(Gly)_7$. $(i+x)$ means contribution of the group i and $i+x$; $x=0$ indicates self contribution, non-zero x indicates the x^{th} neighbor correlation contribution. Slopes of outer term versus n are -2.23 kcal/mol (water), -2.43 kcal/mol (urea), -2.20 kcal/mol (TMAO). All values are in kcal/mol.

	Water	Urea	TMAO
$(i+0)$	-1.53	-1.97	-1.53
$(i+1)$	-0.79	-0.79	-0.79
$(i+2)$	0.25	0.25	0.25
$(i+3)$	-0.11	-0.11	-0.11
Total	-2.18	-2.62	-2.18

We present decomposition of long-range contribution to the outer term in Table 5.2 following Equation, 5.2. Two things are clear from this table: correlation with the neighbors is non-zero in all solvents; correlations with neighbors are the same for all solvents. The formers renders group additivity invalid (because contributions are context, near neighbor, dependent) while the latter suggests why group additivity appears to be valid in peptide transfer free energies. As this simple system is transferred from one solvent to the other, these correlations cancel.

5.3.2.2 Correlations in packing and chemistry

Bayes' rule provides a method for identifying correlations in packing and chemistry (both of these are related to probabilities). If we break the polymer into a distinguished group i and the rest (here denoted as background), then the probability of observing a cavity of the size of the peptide, x_0, is related to the probability of observing a cavity of the size of the distinguished group, $x_0(i)$, and the probability of observing a cavity of the size of the background, $x_0(i_{back})$, by

$x_0 = x_0(i \cdot i_{back}) = x_0(i)x_0(i_{back})x_c$. x_c is the correlation contribution here. x_c can then be evaluated from x_0, $x_0(i)$ and $x_0(i_{back})$. We obtain $x_0(i)$ and $x_0(i_{back})$ using the same method as was used for x_0. Similar, analysis can be performed for packing too.

To demonstrate correlations in packing, we only pursue urea and water. TMAO is left out because of very small difference between packing in TMAO and packing in water. We present, correlations contribution in packing for two different chain lengths in Table 5.3. We take middle of the chain as the distinguished residue and rest as background in both the cases shown in the Table 5.3.

Table 5.3: Correlation contributions in packing. p_0, $p_0(i)$, and $p_0(i_{back})$ mean packing for peptide, packing for distinguished residue and packing for background, respectively. i refers to the distinguished residue which is the middle residue of the chain. A peptide with $n = y$ has $y + 1$ peptide units. No standard error in $-kT \ln p_c$ is greater than 0.4 kcal/mol. All values are in kcal/mol.

solvent	n	p_0	$p_0(i)$	$p_0(i_{back})$	p_c
Water	4	62.0	22.7	60.5	−21.2
	6	81.1	22.9	78.8	−20.6
Urea	4	65.9	24.5	64.2	−22.8
	6	85.7	24.6	83.2	−22.1

Notice that correlation contribution, $-kT \ln\ p_c$, is non-zero for both water and urea. But more strikingly, correlation contribution in water and correlation contribution in urea is not the same! Perhaps the most surprising consequence of this is that correlations in transfer of simple cavities from water to urea do not cancel and additivity does not hold even in the transfer of these simple solutes.

Why, then, does additivity appear to hold? The answer lies in chemistry.

In Table 5.4, we present packing plus chemistry. Presenting packing and chemistry

together makes their comparison easier as we have already seen packing in Table 5.3.

Table 5.4: Correlation in packing+chemistry. x_0/p_0 (packing+chemistry of for the peptide), $x_0(i)/p_0(i)$ (packing plus chemistry for the distinguished residue), and $x_0(i_{back})/p_0(i_{back})$ (packing plus chemistry for the background) are presented in the units of kcal/mole. No standard error in $k_{\mathrm{B}}T \ln x_c/p_c$ is greater than 0.6 kcal/mol. All values are in kcal/mol.

solvent	n	$\frac{x_0}{p_0}$	$\frac{x_0(i)}{p_0(i)}$	$\frac{x_0(i_{back})}{p_0(i_{back})}$	$\frac{x_c}{p_c}$
Water	4	-12.4	3.3	-7.4	-8.3
	6	-17.2	3.2	-12.9	-7.5
Urea	4	-13.0	3.4	-7.9	-8.5
	6	-18.5	3.5	-14.5	-7.5

Observe, in Table 5.4, how the correlation contributions (correlations in packing plus correlations in chemistry) are the same for both urea and water. Thus while additivity does not hold for solvation free energies but it does appear to hold in the transfer free energies. We note that additivity appears to hold in transfer free energies because correlations at two separate energy scales (packing effects are fundamentally different from chemistry) cancel each other.

We saw that there are nontrivial context dependent solvent mediated correlations in solvation free energies that render group additivity invalid. Besides the context dependence of correlations, completely different energy scales have competitive correlations that largely cancel in transfer from one solvent to other. But this cancellation of correlation arises from disparate energy scales, and cautions us against interpreting any mechanism of protein folding with the help of group additivity. These context dependent correlations of binding energy are not limited to the first hydration shell of a polymer, in context dependent correlations persist in simplest of cases such as solvation of hydrophobic cavities and long-range non-specific solute-solvent binding

energies.

After discussing the limitations of group transfer free energy model for the most numerous group of a proteins i.e. the backbone, we now turn our attention to the side-chains. As the hydrophobic effect is often suggested to be the standard explanation for protein folding, we have studied a hydrophobic group (isoleucine) here.

5.4 Hydrophobic side-chain

The hydration thermodynamics of analogs of amino acid sidechains (or of amino acid sidechains in small model peptides) have often been used to understand protein folding and protein-protein association. Tanford [19] formulated this approach into a quantitative, predictive framework. In his approach, the free energy of unfolding is given as a sum of the free energy of transfer of "the small component groups of the molecule, from the environment they have in the native form, to the environment they have in the unfolded form". Accounting for subsequent refinements that included corrections for the solvent-exposure of the "small component groups", equation 5.1 is applied to calculate the unfolding free energy ($\Delta G_{N\to U}$). In interpretation of calorimetric data [116], for example, a gas-phase reference is natural, as is also the case in this section.

At a time when theory, simulations, and experimental techniques were less developed than they are now, the group-additive approach was a pragmatic first step to understand the hydration thermodynamics of a complicated macromolecule. But it is also important to assess its limitations and probe if the physical conclusions based on this approach are valid. For example, the group-additive model appears to capture the effect of osmolytes rather well [152], with the predicted and experimentally determined m-values agreeing to within a couple kcal/mol for proteins with about 100 residues. Further, our earlier study on the solvation of a peptide group [72] showed that in the water-to-osmolyte transfer solvent-mediated correlations be-

tween peptide units (in Gly_n) cancel and this allows one to identify a peptide-group transfer free energy that is model independent. But in the vapor-to-liquid transfer, a situation where solvent-mediated correlations are preserved, the identified group-additive contribution (Δg_i, Eq. 5.1) depends on the model used to define the peptide group. Several recent studies have explored the limitations of the additive model [72, 84, 122, 123]. While our work [72] emphasized solvent-mediated correlations, König et al. [84, 123] have emphasized the role of "self-solvation" [159] in limiting additivity. In exploring context dependence of hydration, these authors also suggested how adding a methyl group to an amino acid changes the solvent-exposure of the peptide backbone thereby influencing the excess free energy in a way that cannot be captured by sidechain analog data. Building on these efforts, here we study how well a model of the sidechain describes the hydration of the sidechain in the context of a protein. In particular, our aim is to better understand factors limiting additivity and the reasons why scaling model-compound data to describe its properties in a protein context may not be satisfactory. To aid in parsing energetic differences, we study the conditional solvation of an isoleucine residue, often regarded as very hydrophobic [160], in the context of model extended and helical peptides. Butane in the *gauche* conformation, matching exactly the sidechain conformation of isoleucine in extended peptides, is used as a small-molecule analog of the sidechain.

Here we use the quasichemical approach to separate the hydrophobic (packing) contributions from the hydrophilic contributions to both facilitate the calculation and also provide insights into the limitations of additivity. These components (but not the net free energy) do depend on the specification of a hydration shell, but the approach has nevertheless provided important insights into the physics of hydration and in generating models of molecular solutions

We complement the quasichemical analysis with the traditional decomposition of the excess free energy into its enthalpic and entropic contributions. The quasichemical approach helps better appreciate how the context influences the hydration of the

solute, while the enthaly-entropy decomposition leads to the finding that non-obvious contributions from the reference (the context) get folded into the conditional contribution attributed to the sidechain (the solute) that is of first interest in additive models.

5.4.1 Methods

The pentapeptides GGGGG, GGIGG, and IGGGG are modeled in the extended configuration with the long axis aligned with the diagonal of the simulation cell and the center of the peptide placed at the center of the simulation cell. Helical deca-glycine (G_9G) and a helical peptide with nine alanine and one glycine at position 6 (A_9G) serve as references for the helical peptides, G_9I and A_9I, respectively. All peptides had an acetylated (ACE) N-terminus and an n-methyl-amide (NME) C-terminus. In the helical peptides, isoleucine was substituted at position 6, approximately in the center of the helix. Butane in the *gauche* conformation was built using the isoleucine conformation in the extended pentapeptide.

The solvent was modeled by the TIP3P [110, 112] model and the CHARMM [114] force field with CMAP correction terms for dihedral angles [113] was used for the peptide. A total of 2006 water molecules solvated the pentapeptide; 3500 water molecules were used for the helical peptides. The Lennard-Jones parameters for the isoleucine sidechain were used for g-butane. Since C_β in isoleucine becomes a CH_2 group in g-butane, the partial charges of that center were slightly adjusted to account for the presence of a capping H-atom. We note that CHARMM does have a parameter set for butane. In particular, the parameter we use for C_β is very slightly shifted from the parameter values for the corresponding carbon atom in butane. Our parameterization was motivated by our desire to be as close to isoleucine as possible, but this minor shift in parameters is not expected to change any conclusions.

The G_9G and A_9G peptides were built from a helical deca-alanine. These struc-

tures were energy minimized with weak restraints on heavy atoms to prevent large distortions of the helix. (After energy minimization, the solute atoms are held fixed for the remainder of the simulation.) The G_9I and A_9I helices were built by grafting the conformation of isoleucine in the GGIGG system onto position 6. Thus the internal conformation of the isoleucine is the same in both g-butane, GGIGG, and the helical peptides.

The free energy calculations and error analysis follow the procedure described in section 2.3 on page 21. Briefly, ϕ_R is applied such that R varies from 0 and 5 Å. For every unit Å, a five-point Gauss-Legendre quadrature rule defines the R points sampled. The work to apply the field is then obtained by quadratures. At each gauss-point, the system was equilibrated for 0.5 ns and data collected over the subsequent 0.5 ns. For the extended peptides, the long-range contribution was obtained by performing particle insertion calculations in the appropriate molecular-shaped cavity. Water with the appropriate cavity was simulated for 1 ns and 1250 frames from the last 0.625 ns used for analysis. For the helical petides, given their high dipole moment we made the conservative choice of obtaining the electrostatic contribution to the long-range interaction using a 2-point Gauss-Legendre quadrature [161]; the van der Waals (vdW) contribution was obtained using particle insertion in the molecular cavity. Electrostatic self-interaction corrections (of about 0.5 kcal/mol) were applied [72]. (For the helices, the sum of the vdW and quadrature-based electrostatic contribution deviates by about 1 kcal/mol from the Gaussian model, but this deviation is significant at the statistical resolution of the chemistry and packing contributions.)

For the zwitterionic isoleucine and glycine we obtained the free energy in two stages. First, the excess free energy of the completely discharged amino acid was obtained using the quasichemical procedure. Then the work required to turn-on the charges was obtained using a 3-point Gauss-Legendre quadrature. The 3-point rule gave the same answer (within 0.1 kcal/mol) as the 2-point rule, but these estimates deviate by over 10 kcal/mol from the linear-response result.

The excess energy was obtained by adapting the shell-wise calculation procedure (section 2.3). For calculating the excess energy we equilibrated the solvated peptide system an additional 1.5 ns (beyond what was used in the free energy calculation), and propagated the trajectory for an additional 3 ns, collecting data every 500 ps for a total of 6000 frames. Entropies obtained using Eq. 2.4 agree within statistical uncertainty with entropy calculated using $-\partial\mu^{\text{ex}}/\partial T$.

5.4.2 Results

Conditional hydration free energy of a molecule XH in the polymer MX is the hydration free energy of MX subtracted by the hydration free energy of MH. Similar definitions are used for conditional entropies and conditional enthalpies of hydration.

While the free energy of hydration is calculated by the regularization scheme, enthalpy hydration is calculated by the shell based scheme. Entropy is then calculated by excess free energy and enthalpy of hydration.

Table 5.5: Conditional thermodynamics of isoleucine in zwitterionic amino acids (indicated by Δ). E_{sw} is the solute-solvent interaction energy and E_{reorg} is the solvent reorganization energy, their summation is h^{ex}. Error bars are drawn at 1σ. E_{reorg}, Ts^{ex} and h^{ex} have approximately equal error bars. All quantities are in kcal/mole.

	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
$G^{(0)}$	1.2 ± 0.1	-4.6 ± 0.5	-8.6 ± 0.02	4.0	-5.8
$I^{(0)}$	2.4 ± 0.1	-5.9 ± 0.5	-14.3 ± 0.02	8.3	-8.3
Δ	1.2 ± 0.1	-1.3 ± 0.7	-5.7 ± 0.03	4.3	-2.5
G	-38.2 ± 0.1	-70.3 ± 1.2	-120.7 ± 0.1	50.4	-32.1
I	-35.7 ± 0.1	-71.2 ± 1.3	-122.5 ± 0.1	51.4	-35.5
Δ	2.5 ± 0.1	-0.9 ± 1.8	-1.8 ± 0.1	1.0	-3.4
g-Butane	2.5 ± 0.1	-3.4 ± 0.5	-9.4 ± 0.02	6.0	-5.9

In Table 5.5, we present conditional hydration free energy of iso-leucine for zwitterionic glycine, subscript 0 indicates a hypothetical amino acid for which partial charges are turned to zero. Notice that the free energy of converting a zwitterionic glycine is the same as the free energy of hydrating a butane. But mutating a hydrogen to butane is easy for a fully discharged amino acid compared to mutating a hydrogen of a fully charged peptide. The reason for this is not the hydration of butane (because exposed surface area of butane is the same in both the cases) but the screening of partial charges of the background by butane. Thus in mutating to an iso-leucine context also goes through a change. This can be seen in the conditional solute-solvent binding energies. When the peptide is discharged the conditional change solute-solvent binding is more negative compared to conditional solute-solvent binding energy when the peptide is fully charged.

Next, we mutate a fully extended pentapeptide GGGGG to GGIGG and IGGGG, the results are shown in Table 5.6. First, notice that even though exposed surface area of butane and background is approximately same, the hydration free energies of GGIGG and IGGGG are different – same surface area does not guarantee same hydration free energies. Secondly, note that not only does the surface area based scaling fail to capture physics of conditional hydration, no single surface area correction is possible that can capture all aspects of the physics.

In Table 5.7, we present the conditional hydration free energies in a helical conformation. We mutate 6^{th} residue of a deca-glycine and another deca-peptide (with a glycine at 6th residue and rest alanines) to isoleucine. This gives us the conditional hydration free energies of butane in the helical conformation for these two peptides. We again find that the surface area scaling can not capture the conditional hydration free energies and the physics of conditional hydration. In fact in this case sign of the conditional solute-solvent binding energies is also estimated wrong.

We calculated conditional hydration free energies of butane and compared it with the prediction from surface area based scaling of hydration free energy of butane. We

Table 5.6: Thermodynamics of mutation of GGGGG. E_{sw} is the solute-solvent interaction energy and E_{reorg} is the solvent reorganization energy, their summation is $h^{\rm ex}$. Fractional exposure of area in butane as isoleucine side chain in GGIGG is $\alpha_{sc} = 0.619$ in GGIGG to that for g-butane; for IGGGG this fraction $\alpha_{sc} = 0.616$. Standard error of the mean is given at 1σ. Error bars are drawn at 1σ. E_{reorg}, $Ts^{\rm ex}$ and $h^{\rm ex}$ have approximately equal error bars. All quantities are in kcal/mole.

	Chem	Pack	LR	$\mu^{\rm ex}$	$h^{\rm ex}$	E_{sw}	E_{reorg}	$Ts^{\rm ex}$
g-Butane	−16.4	24.0	−5.1	2.5 ± 0.1	-3.4 ± 0.5	−9.4	6.0	−5.9
GGGGG	−85.8	71.1	−17.3	-32.0 ± 0.3	-56.9 ± 1.4	−105.5	48.6	−24.8
GGIGG	−88.2	76.8	−17.7	-29.1 ± 0.3	-56.3 ± 1.4	−107.3	51.0	−27.3
IGGGG	−88.0	76.3	−18.3	-30.0 ± 0.3	-57.3 ± 1.8	−107.6	50.3	−27.3
Δ[GGIGG-GGGGG]	−2.4	5.7	−0.4	2.9 ± 0.4	0.6 ± 2.0	−1.8	2.4	−2.5
α_{sc}·[g-butane]	−10.2	14.9	−3.2	1.5	−2.1	−5.8	3.7	−3.7
Δ[IGGGG-GGGGG]	−2.2	5.2	−0.9	2.1 ± 0.4	-0.5 ± 2.0	−2.1	1.7	−2.6
α_{sc}·[g-butane]	−10.1	14.8	−3.1	1.5	−2.1	−5.8	3.7	−3.7

find that the conditional hydration of butane also changes the context; again, no single area scaling factor can capture all the aspects of hydration thermodynamics of the proteins. When a glycine is mutated to iso-leucine the binding energy distribution of the background also goes through a change because of this mutation. Thus caution must be taken in interpreting physical mechanism of protein folding using group transfer free energies.

5.5 Summary

We have analyzed the group transfer free energy model for the solvation thermodynamics of a hydrophilic group (the backbone) and a hydrophilic group (butane-*analog*). We find that this assumption has serious limitations in interpreting the thermodynamics of protein solvation.

Table 5.7: Conditional thermodynamics of isoleucine in deca-peptides. All peptides are in helical conformation

	Chem	Pack	LR	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
G_9G	−95.1	78.0	−32.2	-49.3 ± 0.8	-83.6 ± 1.6	−153.6	70.0	−34.3
G_9I	−97.3	83.7	−32.0	-45.6 ± 0.8	-82.5 ± 2.2	−155.9	73.4	−36.9
Δ[G_9I - G_9G]	−2.2	5.7	0.2	3.7 ± 1.1	1.1 ± 2.7	−2.3	3.4	−2.6
α_{sc}·[g-butane]	−9.7	14.2	−3.0	1.5	−2.0	−5.5	3.5	−3.5
α_{sc}·[GGIGG − GGGGG]	−2.3	5.4	−0.4	2.7	0.6	−1.7	2.3	−2.4
A_9G	−95.1	86.7	−30.2	-38.6 ± 0.9	-78.1 ± 3.0	−152.1	74.0	−39.5
A_9I	−95.8	90.7	−30.3	-35.4 ± 0.8	-74.8 ± 2.5	−151.1	76.3	−39.4
Δ[A_9I - A_9G]	−0.7	4.0	−0.1	3.2 ± 1.2	3.3 ± 4.0	1.0	2.3	0.1
α_{sc}·[g-butane]	−9.2	13.4	−2.9	1.3	−1.9	−5.3	3.4	−3.3
α_{sc}·[GGIGG−GGGGG]	−2.2	5.1	−0.4	2.5	0.5	−1.6	2.2	−2.3

Group transfer free energy model assumes that each residue interacts independently (in the absence of correlations between two any residues) with the solvent. This assumption is violated in all the cases studies here – neighboring residues do influence solvent binding with each other. We have found nontrivial near neighbor solvent mediated correlations between two residues of protein in this work. These correlations make the group additivity based assumptions unsuitable to interpret solvation thermodynamics.

6

Helix-coil transition in deca-peptide models

Note: This chapter draws upon a manuscript that is currently being circulated to other researchers for comments prior to submission to a peer reviewed journal.

6.1 Overview

In chapter 3 and chapter 4 we showed that thermodynamics of protein hydration is different from widely accepted explanations. We found in chapter 5 that those earlier explanations were based on an assumption that is ill suited to interpret protein hydration thermodynamics. In this chapter, we will take a prototypical case of helix-coil transition and study it using quasichemical theory to better understand mechanism of protein folding.

We calculate the hydration free energy of a deca-alanine peptide in the α-helix and several coil states using a quasichemical organization of the potential distribution theorem. Hydrophobic (packing) effects favor the compact folded states, as expected. Surprisingly, this effect is outweighed by attractive protein-water interactions that favor the coil states: hydration favors the unfolded state, suggesting that protein

intramolecular interactions are decisive in forming the helix. In the pairing of two helices, packing outweighs short-range attractive protein-water interactions and drives helix pairing. However, long-range attractive protein-water interactions, that are sensitive to the relative orientation of the helices, can further enhance or reverse this trend. The decomposition of free energies into entropic and enthalpic components shows that change in entropy drives helix-pairing, but change in enthalpy, primarily arising from changes in attractive protein-solvent interactions, opposes helix pairing. In helix-pairing as well, changes in protein-protein interaction energy are found to be as important as solvent effects.

6.2 Introduction

The α-helix is a fundamental secondary structural element in folded proteins, making the understanding of forces stabilizing this structure of pre-eminent interest in the broader quest to understand the thermodynamic driving forces underlying protein folding [121]. Within the currently available molecular mechanics approaches it is fairly straightforward to assess the role of peptide-intramolecular interactions in the formation of the secondary structures. However, assessing the role of the solvent in this process still remains a challenge.

Here, we use the regularization approach to free energy calculations to study hydration effects in the classical problem of the coil-to-helix transition in a deca-alanine peptide. In addition, we also study the pairing of two helices as a simplified model of secondary-to-tertiary assembly. Our models of primary-to-secondary and secondary-to-tertiary folding also allows us to investigate how the thermodynamic driving forces originating in hydration effects behave at different length scales.

We find that hydrophobicity is not consequential in the formation of a secondary structure, but it does play a substantial role in the helix-helix complexation. Long-range interactions that have traditionally been regarded as unimportant in the ther-

modynamics of folding, especially for proteins at the isoelectric point [32], are shown to play a nontrivial role in protein hydration, coil-to-helix folding, and helix-helix assembly. Finally, changes involving peptide intramolecular interactions are also important, both in folding and in assembly. Importantly, we find that no single contribution can be ignored in assessing the thermodynamics of folding.

Experimental studies of solvent effects at the scale of a protein, for example in the molecular interpretation of calorimetric data [116], typically appeal to group-additive approaches that simulations [72, 73, 84, 122, 123] now reveal to have significant limitations. Simulation studies seeking hydration thermodynamics — free energy of hydration and its derivatives — at the scale of a protein still largely appeal to coarse-grained descriptions of the solvent [52]. Using such methods, including lattice models of protein folding that implicitly account for solvent effects in the potentials, past studies have come to substantially differing conclusions about solvent effects in the coil-to-helix transition: some have suggested that hydrophobicity drives folding [41, 124], while others have emphasized the role of favorable electrostatics [42].

We will next present our simulation methods followed by results.

6.3 Methods

In addition to section 2.3 on page 21, the following completes the methods for this chapter.

The deca-alanine peptide was modeled with an acetylated (ACE) N-terminus and n-methyl-amide (NME) capped C-terminus. The extended β-conformation ($\phi, \psi = -154 \pm 12, 149 \pm 9$) was aligned such that the end-to-end vector lay along the diagonal of the simulation cell. We label this coil state as C_0. The helix was aligned with its axis along the x-axis of the cell. The initial structures were energy minimized with weak restraints on the heavy atoms to relieve any strain in the structure. The peptides were solvated in 3500 TIP3P [110, 112] water molecules. Version c31 of the

CHARMM [114] forcefield with correction terms for dihedral angles [113], was used for the peptides. We also compute the free energy of unfolding the protein in vacuum using the adaptive-bias force (ABF) approach [125, 126]. From this trajectory, we sampled nine structures with end-to-end distances between terminal carbon atoms ranging between 28 Å and 36 Å in increments of 1 Å. We label the coil states from this unfolding simulation $\{C_1, \ldots, C_9\}$. The ϕ, ψ for these unfolded states predominantly populate β and PPII regions of the Ramachandran plot. Structures were held fixed in the hydration studies.

The same set-up is used to investigate helix pairing. We consider the potential of mean force (PMF), $W(r)$ along, r, the separation between the helix axis with the axis parallel to each other. Additionally, we consider two relative orientations of the helix dipoles, parallel and antiparallel. (Note that the helix dipoles will be antiparallel in the simplest helix-turn-helix motif.) These arrangements help illuminate the role of long-range protein-solvent interactions in helix-helix complexation. The PMF

$$W(r) = W_{solv}(r) + \Delta U(r)\,, \tag{6.1}$$

where W_{solv} is the solvent (or indirect) contribution [71] and ΔU is the contribution from direct protein-protein interactions relative to the proteins infinitely apart. $W_{solv}(r) = \mu^{\text{ex}}(r) - 2\mu^{\text{ex}}$, where $\mu^{\text{ex}}(r)$ is the hydration free energy of the pair of helices (for a given separation and orientation) and μ^{ex} is the hydration free energy of an individual helix. We calculated ΔU using the same forcefield as that used in the dynamics.

We also sought the enthalpic and entropic decomposition of $W_{solv}(r)$ using the shell-based scheme (section 2.3).

Table 6.1: Components of the hydration free energy for the helix and the least favorably (C_0) and most favorably (C_7) hydrated coil states. For the helix and C_0 states, results with partial charges turned off (indicated by $\mathbf{Q} = \mathbf{0}$) are also noted. R_g is the radius of gyration (relative to the center of mass) and R_c is the end-to-end distance between terminal carbon atoms in Å. SASA is the solvent accessible surface area in Å^2. All energy values are in kcal/mol. Standard error of the mean is given in parenthesis at the 2σ level.

Conformation	R_g	R_c	SASA	Packing	Chemistry	Long-range	μ^{ex}
Helix	5.3	16.3	876.0	87.4 (0.4)	−94.6 (0.2)	−31.6 (0.2)	−38.8 (0.5)
Helix ($\mathbf{Q} = \mathbf{0}$)					−65.4 (0.3)	−18.1 (0.03)	3.9 (0.5)
Coil (C_0)	11.1	36.8	1260.0	127.8 (0.3)	−146.5 (0.3)	−27.6 (0.1)	−46.3 (0.4)
Coil ($\mathbf{Q} = \mathbf{0}$)					−100.3 (0.4)	−24.7 (0.03)	2.8 (0.5)
Coil (C_7)	10.6	34.0	1249.0	127.2 (0.5)	−152.3 (0.3)	−28.3 (0.4)	−53.4 (0.8)

6.4 Results

6.4.1 Coil-to-helix transition

In Table 6.1 we collect the results of the free energy calculation on the helix state, the C_0 coil state and the most favorably hydrated of the coil states in $\{C_1, \ldots, C_9\}$. (The least favorably hydrated coil state is C_0 itself.)

Table 6.1 indicates that at least for the coil states considered here, the packing is somewhat insensitive to the peptide structure. At the scale of the cavities, we expect the packing contributions to scale with surface area [127–129] and the data in the table conforms to this expectation. Since the solvent-accessible surface areas of C_0 and C_7 are not very different, despite overall differences in the structure, we expect the packing contribution to be similar for these states, as found in simulations. The packing contribution favors the helix state by about −40 kcal/mol: hydrophobic hydration greatly favors the compact state of the protein. The chemistry contribution

is, however, in the opposite direction. It favors the C_7 coil state by -57.7 kcal/mol and the C_0 state by -51.9 kcal/mol. Thus the local protein-solvent interaction outweighs the packing contribution by between 11 kcal/mol to 18 kcal/mol in favoring the coil state, a conclusion that is invariant to the choice of physically realistic R (Table 6.2). These favorable local protein solvent interactions arise primarily from favorable peptide backbone (CONHCα)-water interactions (Table S.I); the $\mathbf{Q} = \mathbf{0}$ results emphasize that both dispersion and electrostatic effects are important for assessing backbone-water interactions.

Table 6.2: Components of the hydration free energy for the helix, C_0, and C_7 states for $R = 3$ Å. For ease of comparison, the $R = 5$ Å data is also provided. All energy values are in kcal/mol. Standard error of the mean is given in parenthesis at the 2σ level.

Conformation	R	Packing	Chemistry	Long-range	μ^{ex}
Helical	5 Å	87.4 (0.4)	-94.6 (0.2)	-31.6 (0.2)	-38.8 (0.5)
	3 Å	44.4 (0.4)	-0.4	-82.8 (0.7)	-38.8 (0.5)
Coil (C_0)	5 Å	127.8 (0.3)	-146.5 (0.3)	-27.6 (0.1)	-46.3 (0.4)
	3 Å	58.4 (0.4)	-0.8	-103.9 (0.4)	-46.3 (0.4)
Coil (C_7)	5 Å	127.2 (0.5)	-152.3 (0.3)	-28.3 (0.4)	-53.4 (0.8)
	3 Å	58.3 (0.4)	-0.8	-110.9 (0.9)	-53.4 (0.8)

Table 6.1 also highlights that attractive solute-water dispersion interactions can drive helix-to-coil transition in a peptide that lacks partial atomic charges. This suggests that we must reevaluate the relevance of the poor solubility of nonpolar solutes — observe that both the helix and coil have a positive μ^{ex} — in rationalizing the collapse of a polypeptide. Our observation that hydration does not explain the collapse of a nonpolar chain is consistent with the observation of similar behavior in alkanes (cf. Ref. [130] and the reanalysis of data in Ref. [131] presented therein). Interest-

ingly, attractive solute-water interactions also oppose the pairing of the prototypical hydrophobe methane [71].

Table 6.3: Decomposition of long-range contribution to the hydration free energy into its van der Waals (*vdw*) and electrostatic components (*elec*). The electrostatic contribution includes self-interaction corrections. All energy values are in kcal/mol.

Conformation	Long-range	*vdw*	*elec*
Helix	−31.6	−18.1	−13.5
Coil (C_0)	−27.6	−24.7	−2.9
Coil (C_7)	−28.3	−24.2	−4.1

Protein-solvent long-range interactions contribute a substantial fraction — roughly between between 50% and 82% — of the net hydration free energy of the peptide. In the coil-to-helix transition, the long-range contributions favor the helix by about −4.0 kcal/mol (Table 6.1), primarily driven by changes in electrostatic interactions, as might be inferred given the high dipole moment of the helix (Table 6.3). The totality of all these hydration contributions still favors the coil state by between 7 and 14 kcal/mol. Thus the net attractive peptide-water interactions outweigh the hydrophobic driving force and favors the *unfolded* state of the peptide.

Past simulation studies on the helix-coil transition have typically interpreted the results within the framework of the Zimm-Bragg or Lifson-Roig formalisms [118, 119, 132, 133]. These studies have provided insights into the enthalpic and entropic contributions to helix formation as well as reveal limitations in forcefields. However, all-atom simulations parsing hydration contributions in the helix-coil transition appear to be scarce. A recent study found that hydration favors the coil state by about 19 kcal/mol [98], qualitatively consistent with this work. (The structure for the coil and the forcefield parameters used are different from that here.) In that study, the free energies were decomposed into electrostatic and van der Waals contributions, in

contrast to the hydrophilic and hydrophobic partitioning sought here.

Ooi and Oobatake [134] have used an extensively parameterized surface area model to compute hydration effects in protein folding. For the coil-to-helix transition in a deca-alanine, with the coil modeled as a fully extended peptide, they found the change in hydration free energy to be about 10 kcal/mol, a value bracketed by our calculations. They find that in the gas-phase, the coil-to-helix folding free energy is −9 kcal/mol. Thus coil is predicted to be more stable in solution, consistent with experiments with similar short peptides [135–139].

The ABF-based estimate of the free energy change for the coil-to-helix transition in the gas-phase is about −29 kcal/mol (Fig. 6.1).

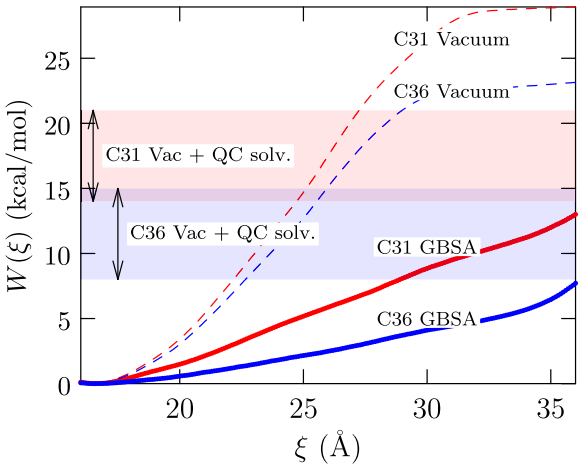


Figure 6.1: The C_0 ($\xi = 36.8$ Å, $\Delta\mu^{\text{ex}} \approx -7$ kcal/mol) and C_7 ($\xi = 34.0$ Å, $\Delta\mu^{\text{ex}} \approx -14$ kcal/mol) coil states have roughly the same $W(\xi)$; $\Delta\mu^{\text{ex}}$ are the all-atom hydration free energies ("QC solv") relative to the helix. The shaded regions correspond to the range of unfolding free energies using an *a posteriori* correction for hydration of the vacuum unfolding free energy. Self-consistent hydration using GB/SA suggests a marginally lower unfolding free energy. Notice that the c36 force-field improves the balance towards the coil state.

Adding correction due to hydration *a posteriori*, the coil-to-helix transition in

water is still predicted to be strongly favored by between $-15 = (14 - 29)$ kcal/mol and $-22 = (7 - 29)$ kcal/mol. (The ABF-based estimate of the coil-to-helix free energy change with self-consistent account of hydration, albeit with the continuum GB/SA model [140–142] of the solvent, gives a value of about -13 kcal/mol.) The corresponding range is -16 to -6 kcal/mol with the recently updated c36 version [143] of the CHARMM forcefield and a value of about -8 kcal/mol with the GB/SA model. (The c36 version improves the helix-to-coil balance relative to c31, as expected based on the aim in designing this forcefield.)

Based on experiments on similar short peptides [135–139], such a strong drive to form a helix in deca-alanine is not suggested. This suggests that the forcefield model is deficient and biased towards the helical state, consistent with the earlier observation by Best and Hummer [144] based on their analysis of NMR J-coupling. Unfortunately, based on our results we cannot infer which aspect of the forcefield is deficient. Based on our limited comparison between C31 and C36 variants of the CHARMM forcefield and earlier studies (cf. [119, 144]) that show that minor tuning of protein dihedral potential can improve agreement with experiments, we suspect that the strong drive to a helix likely reflects problems with the intra-molecular energy function. Further, the nonbonded solute-solvent interactions have typically been parametrized and extensively tested against hydration thermodynamics of small molecules [145] and this part of the forcefield is thus expected to be more reasonable.

We expect the role of hydration in unfolding the helix to hold for larger chain-lengths, provided the coil states are such that the backbone remains accessible to solvent. Since coil-to-helix transition is seen in peptides with as few as 13 residues [137], assuming the solute-solvent interaction model is reasonable, our results suggest that folding of the helix is driven primarily by changes in protein intra-molecular interactions, an inference that is in consonance with enthalpy changes driving helix formation [136].

6.4.2 Helix-helix complexation

We next consider the potential of mean force between two helices. Fig. 6.2 shows

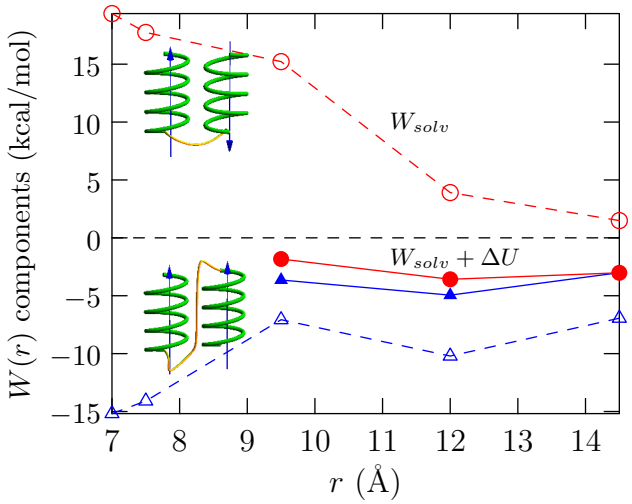


Figure 6.2: Components of the potential of mean force in bringing two helices together. The helices are shown in green and the connecting yellow tube is meant to suggest one way these helices could be organized along a linear polypeptide. The parallel (△) and antiparallel (◯) arrangements are also highlighted by the blue-arrow denoting the helix dipole. W_{solv} is the solvent contribution (open symbols), and $W_{solv} + \Delta U$ (Eq. 6.1) is the net PMF (filled symbols). For $r \lesssim 8$ Å, there is steric overlap between the helices and ΔU rises rather sharply. Data including these values of ΔU are thus not shown.

that hydration opposes the complexation of helices in the antiparallel orientation. Our results thus suggest that hydration would oppose formation of the commonly found helix-turn-helix motif. Interestingly, the direct contributions (ΔU) can outweigh the hydration effects to drive complexation. (These direct contributions involve non-bonded electrostatics and van der Waals contributions; hydrogen bonding effects that are commonly invoked in folding is subsumed in these effects.) In the antiparallel arrangement favorable ΔU drives complexation, whereas for the parallel

arrangement, unfavorable ΔU tempers the favorable hydration effects. The net free energy of complexation is roughly -2 kcal/mol, a small magnitude relative to the large competing hydration and inter-molecular interaction effects.

Parsing the hydration contributions shows that in contrast to the coil-to-helix transition hydrophobic hydration drives helix-helix complexation (Fig. 6.3). But long-

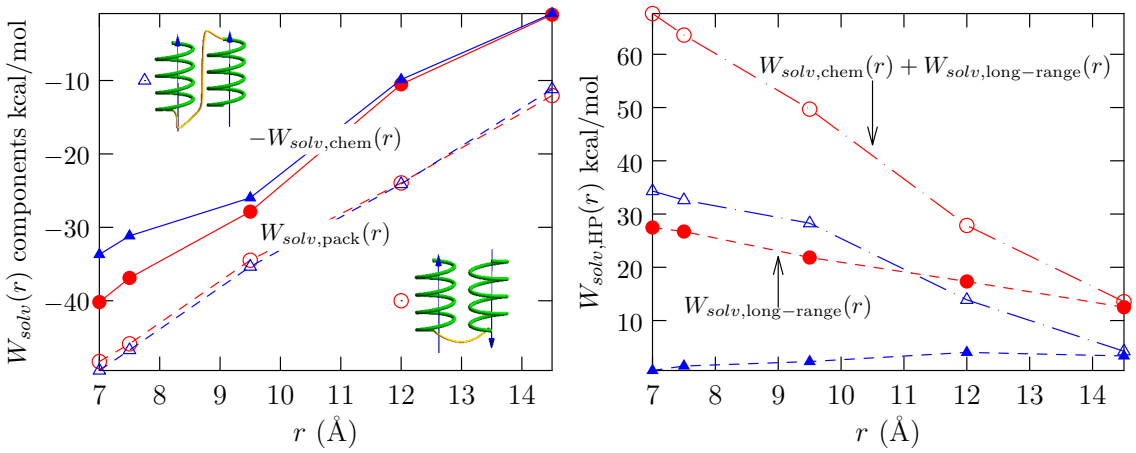


Figure 6.3: Left panel: The packing (open symbols) and chemistry (filled symbols) contributions to the potential of mean force. The negative of the chemistry contribution is shown to aid in direct comparison with the packing contribution. Right panel: the hydrophilic contribution (open symbols) $W_{\mathrm{HP}} = W_{\mathrm{chem}} + W_{\mathrm{pack}}$ to the potential of mean force. The long range contribution (filled symbols) is also shown separately.

range interactions play an entirely nontrivial role in the complexation process. The antiparallel arrangement of the helices is strongly disfavored by loss of favorable long-range solute-solvent interactions in hydration, while the inhibition is more modest for the parallel arrangement of helices. Importantly, irrespective of helix orientation, the net hydrophilic contribution given by the sum of the chemistry and long-range contributions disfavors helix pairing (Fig. 6.3, right panel); as expected, this conclusion is insensitive to the choice of a meaningful R. Parsing the long-range interaction into van der Waals and electrostatic components shows that this inhibition is solely due to electrostatic interactions (Figure 6.4).

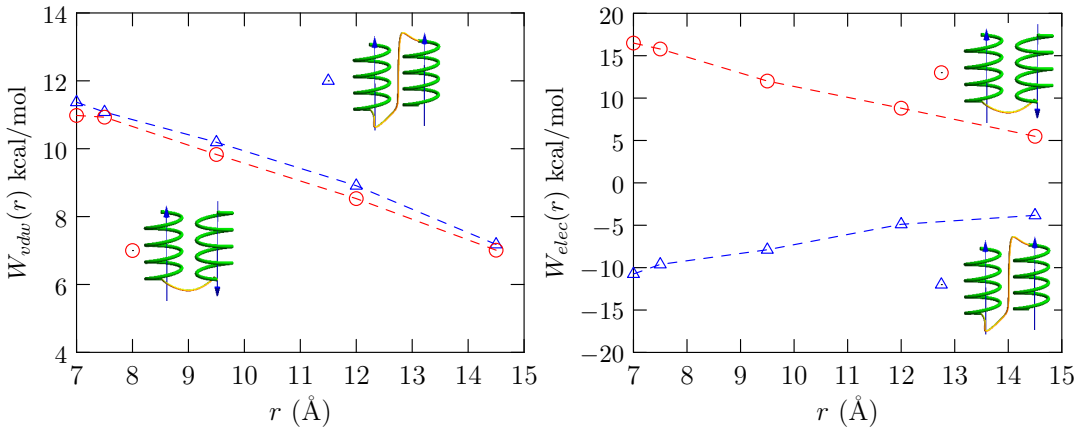


Figure 6.4: Long-range van der Waals (left) and electrostatic (right) contributions in helix pairing.

Parsing W_{solv} into its enthalpic and entropic components also proves illuminating (Table 6.4). The change in the hydration entropy favors helix pairing for both the configurations, but the characteristics of the change in reorganization and interaction components emphasizes the need for caution in interpreting this change in terms of specific changes in water structure. (For $r = 14.5$ Å, the excess entropy is negative, but the reorganization contribution is positive, an observation that is not readily rationalized within the classic iceberg picture of hydrophobic hydration.) The change in hydration enthalpy disfavors helix pairing for either arrangement of the helices. For the antiparallel configuration, the change in the water reorganization energy favors helix pairing. But this change is overwhelmed by the loss of backbone-solvent and sidechain-solvent interactions, that is the hydrophilic contributions. For the parallel arrangement, the change in the water reorganization contribution is approximately zero, but helix pairing is once again inhibited by the solute-solvent interaction contributions. Emphasizing the importance of electrostatic interactions between the backbone and the solvent, the backbone-solvent contributions are sensitive to the orientation of the helices, but the sidechain-solvent contributions are similar.

Table 6.4: Enthalpic (h^{ex}) and entropic contributions ($T\Delta s^{\text{ex}}$) to the change in the hydration contribution to the potential of mean force, W_{solv}, as the helices are brought from 14.5 Å to 9.5 Å. The change in hydration enthalpy, Δh^{ex}, is further divided into ΔE_{reorg}, the change in the water reorganization contribution, and ΔE_{sw}, the peptide water interaction contribution. The latter is subdivided into contributions from the backbone-water interactions, ΔE_{bb}, and sidechain-water interactions, ΔE_{sc}. $\Delta X = X(9.5\text{ Å}) - X(14.5\text{ Å})$, for all X. Standard error of the mean is given in parenthesis at the 1σ level. The value of $Ts^{\text{ex}}(r = 14.5\text{ Å})$ is explicitly given to emphasize that the entropy of hydration is negative; likewise $E_{reorg}(r = 14.5\text{ Å})$ is positive.

Conformation	ΔE_{bb}	ΔE_{sc}	ΔE_{reorg}	Δh^{ex}	ΔW_{solv}	$T\Delta s^{\text{ex}}$
Antiparallel	25.8 (0.5)	5.1 (0.1)	−7.5 (5.0)	23.4 (5.0)	13.8 (1.2)	9.6(−70.2)
Parallel	1.7 (0.4)	6.7 (0.1)	0.8 (5.0)	9.2 (5.0)	−0.1 (1.2)	9.2(−69.3)

6.5 Summary

Our study leads to several important findings. One, hydrophilic hydration opposes both coil-to-helix folding and secondary-to-tertiary transition, here modeled by the pairing of two helices. The latter model does ignore the role the loops connecting the helices play in the pairing, but it does suggest itself as a reasonable starting point to understand solvent effects in tertiary structure formation.

Two, hydrophobic effects tend to favor the compact state of the polypetide and favor both the coil-to-helix transition and helix-helix complexation. But in the coil-to-helix transition, hydrophilic effects (protein-water attractive interactions) easily overwhelm the hydrophobic contribution and favor unfolding of the peptide. Even for a discharged peptide, essentially a nonpolar chain, attractive solute-solvent dispersion interactions suffice to favor the helix-to-coil transition.

Three, in the pairing of helices, a phenomena occurring at a larger length-scale

than the coil-to-helix transition of a single peptide, hydrophobic effects do outcompete the short-range peptide-water interactions in favoring helix complexation. But in this case, too, the long-range protein-solvent attractive interactions, especially for the antiparallel arrangement of helices, outweighs the net effect of the packing plus short-range attraction contributions to favor the disassembly of the helices.

Four, in both the coil-to-helix transition and in the pairing of the helices in the antiparallel orientation of the helix dipoles, we find that changes in the intra-molecular energy of the protein is essential in shifting the balance to the folded state. This intra-molecular interaction is primarily due to van der Waals and electrostatic interactions. (Within the forcefield, all effects of hydrogen-bonding are subsumed in these interactions.)

The limitations of the models and forcefield notwithstanding, our study suggests that in protein folding hydrophilic effects and protein intra-molecular interactions are as important as, if not more important than, hydrophobic effects.

7

Conclusions and future directions

In the study of unfolding $\rightleftharpoons$ folding transition of proteins, experimentalists have relied upon small molecule data while simulation studies have analyzed the thermodynamics of hydration (excess free energy, entropy, and enthalpy) for objects as large as proteins, only in the cases where the solvent is treated as a featureless continuum. We are not aware of any theoretical study that calculates thermodynamics (free energy) of protein solvation while doing justice to the atomistic nature of the solute and the solvent. The achievement of this thesis has been to make these calculations routine.

With the help of these calculations, we have found, hydrophilic effects dominate the hydrophobic effect in the formation of secondary structures; for the tertiary structure formation the hydrophobic effect does play a substantial role but at that length-scale physics of hydrophobic effects is governed by the surface tension of the solvent, not by the conventional picture of "hydrophobic bond" suggested by Kauzmann.

In our study on the deca-alanine peptide we have found all the signatures of hydrophobicity (Chapter 3), but further analysis of these signatures led us to conclude that their origin is in the hydrophilic effects. This suggests a fundamental flaw in our interpretation of the hydration free energies. In fact, in another study, we had found that the common method used to interpret hydration free energies (the group

additivity based scheme) has some limitations (Chapter 5). This thesis suggests that caution must be exercised in translating small molecule data to a molecule as complex as protein; in a protein, nontrivial correlations between the sidechain, backbone, and solvent do not allow a perfect translation of small molecule data.

For osmolyte effect on protein solubility (Chapter 4), our work finds that solute-solvent attractive interactions are more important in protein solubility than the hydrophobic effect. Our work supports a broader range of hydrophilic effects. In the role of long-range, dispersion interaction plays significant role in transfer free energies and m-value. In the short-range, we observed TMAO acts via a backbone based mechanism while urea acts via promiscuous attractive interactions with the protein. Moreover, the important role of dispersion interactions becomes clear in the case of guanidinium hydrochloride as it unfolds proteins with the help of dispersion interactions.

We later treated the helix-coil transition using our framework (Chapter 6). This study leads to several important findings. One, hydrophilic hydration opposes both coil-to-helix folding and secondary-to-tertiary transition, here modeled by the pairing of two helices. Two, hydrophobic effects tend to favor the compact state of the polypetide and favor both the coil-to-helix transition and helix-helix complexation. But in the coil-to-helix transition, hydrophilic effects (protein-water attractive interactions) easily overwhelm the hydrophobic contribution and favor unfolding of the peptide. Even for a discharged peptide, essentially a nonpolar chain, attractive solute-solvent dispersion interactions suffice to favor the helix-to-coil transition. Three, in the pairing of helices, a phenomena occurring at a larger length-scale than the coil-to-helix transition of a single peptide, hydrophobic effects do out-compete the short-range peptide-water interactions in favoring helix complexation. But in this case, too, the long-range protein-solvent attractive interactions, especially for the antiparallel arrangement of helices, outweighs the net effect of the packing plus short-range attraction contributions to favor the disassembly of the helices. Four, in

both the coil-to-helix transition and in the pairing of the helices in the antiparallel orientation of the helix dipoles, we find that changes in the intra-molecular energy of the protein is essential in shifting the balance to the folded state. This intra-molecular interaction is primarily due to van der Waals and electrostatic interactions.

The limitations of the models and forcefield notwithstanding, **this thesis suggests that in protein folding, hydrophilic effects and protein intra-molecular interactions are as important as, if not more important than, hydrophobic effects.**

7.1 Future directions

7.1.1 Protein conformational switches

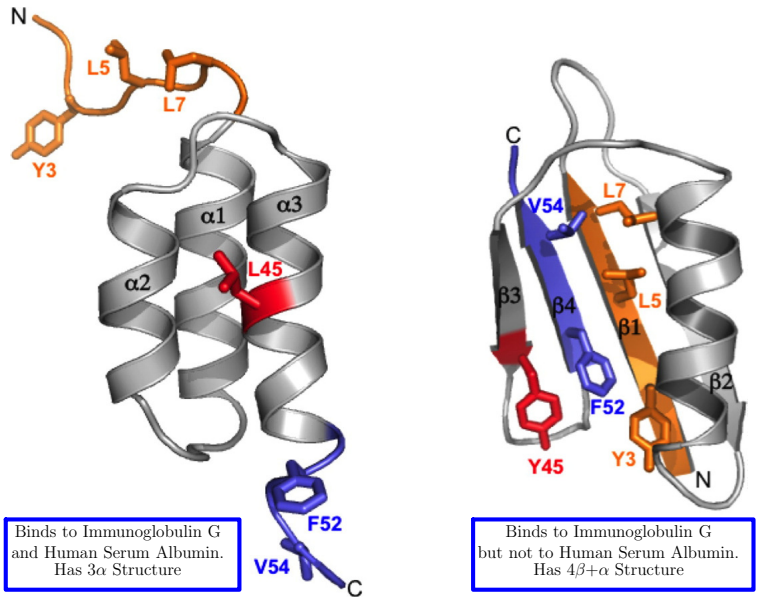


Figure 7.1: Proteins shown above have different functions and fold, yet 44 of their 45 residues are the same. Picture taken and modified from reference [162]

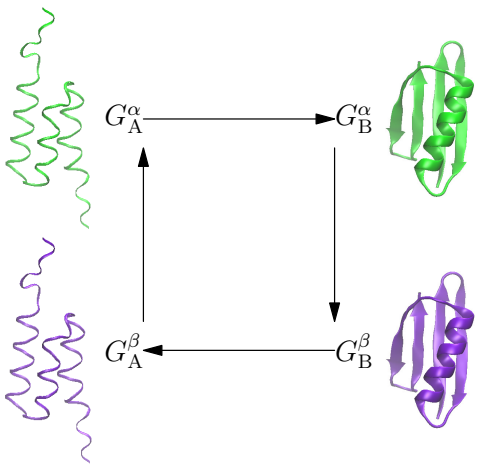


Figure 7.2: A simple cycle to understand the effect of mutation on the structural stability. Each arrow represents free energy contribution in that step. Quasichemical theory will provide hydrophobic and hydrophilic contribution in each step.

A recent study [162] on peptide conformational switching has shown that a mutation in just one out of 45 amino-acids of a sequence can lead to dramatic changes in fold and function (Figure 7.1). Such conformational switches (involving small number of mutations) are appealing candidates to understand the role of solvent in the the sequence-structure relationship as a small change in the amino-acid sequence (thus small change in the protein's interactions with itself) can bring a noticeable difference in the hydration structure of the protein. It is at this point the framework developed in this thesis can be helpful, as no other presently existing techniques can calculate hydration free energies of proteins with the kind of atomistic details needed to understand the role of solvent in protein conformational switches.

As a starting step, we suggest a simple cycle (akin to Hess's law) involving hydration free energies. For the purpose of illustration we will denote the proteins by $\mathrm{G}^{\alpha}_{\mathrm{A}}$, $\mathrm{G}^{\alpha}_{\mathrm{B}}$, $\mathrm{G}^{\beta}_{\mathrm{A}}$, $\mathrm{G}^{\beta}_{\mathrm{B}}$; the subscript indicate fold and superscripts indicate sequence. In the study in reference [162], $\mathrm{G}^{\alpha}_{\mathrm{A}}$, $\mathrm{G}^{\beta}_{\mathrm{B}}$, are stable proteins, while $\mathrm{G}^{\alpha}_{\mathrm{B}}$, $\mathrm{G}^{\beta}_{\mathrm{A}}$ are unstable. The cycle in the spirit of quasichemical theory (role of hydrophobic and hydrophilic effects in each step of the cycle) can be drawn as Figure 7.2. In this cycle hydration

free energies of G_A^α, G_B^α, G_A^β, G_B^β are required to calculate the role of solvent. We have in fact, as a test, already calculated hydration free energies of a few of these structures with our framework.

7.1.2 Osmolytes and quasichemical theory

Here we will show that our framework can be further developed to transparently understand the role of osmolytes in protein hydration thermodynamics. In our framework, chemical potential is written as summation of packing, chemistry and long-range contribution. Because long-range contribution is a result of well behaved gaussian distribution, the role of osmolyte and water can be easily segregated in the long-range. Here, we will discuss the chemistry contribution. (These equation can also be written for packing contribution. As for a big enough cavity size the physics of packing contribution approaches the physics of surface tension, these ideas can also be extended to understand the role of surfactants in changing surface tension with the help of experimental data.)

Let there be a molecular solute, P, in pure water. Following quasichemical theory [60], The chemistry contribution in pure water is given by the following equations

$$PW_n + i\,W \leftrightharpoons PW_{n+i} \tag{7.1}$$

$$K_n \rho^n = \frac{c_n}{c_0} \tag{7.2}$$

$$\underline{x}(W=0) = \frac{1}{1 + \sum\limits_{n\geq 1} k_n \rho^n} \tag{7.3}$$

We adopt a notation in which ρ is the concentration of pure water, $\underline{x}(W=n)$ is the probability of observing n water molecules in the cavity around the solute, and K_n is the equilibrium constant in pure water.

If this molecular solute is transferred to a solution of water and a co-solute (denoted

here by U.) Then we have

$$x(W=0,U=0)=\frac{1}{1+\sum\limits_{n\geq 1}k_{n0}\rho_w^n+\sum\limits_{m\geq 1}k_{0m}\rho_U^m+\sum\sum\limits_{n\geq 1,m\geq 1}k_{nm}\rho_w^n\rho_U^m} \tag{7.4}$$

We also have

$$x(W=n,U=m)=\frac{k_{nm}\rho_w^n\rho_U^m}{1+\sum\limits_{n\geq 1}k_{n0}\rho_w^n+\sum\limits_{m\geq 1}k_{0m}\rho_U^m+\sum\sum\limits_{n\geq 1,m\geq 1}k_{nm}\rho_w^n\rho_U^m} \tag{7.5}$$

$$x(W=0|U=0)=\frac{x(W=0,U=0)}{x(U=0)}=\frac{1}{1+\sum\limits_{n\geq 1}k_{n0}\rho_W^n} \tag{7.6}$$

$$x(W=n|U=0)=\frac{x(W=n,U=0)}{x(U=0)}=\frac{k_{n0}\rho_w^n}{1+\sum\limits_{n\geq 1}k_{n0}\rho_W^n} \tag{7.7}$$

$$x_{*0}=\frac{1+\sum\limits_{n\geq 1}k_{n0}\rho_W^n}{1+\sum\limits_{n\geq 1}k_{n0}\rho_W^n+\sum\limits_{m\geq 1}k_{0m}\rho_U^m+\sum\sum\limits_{n\geq 1,m\geq 1}k_{nm}\rho_w^n\rho_U^m} \tag{7.8}$$

$$x_{0*}=\frac{1+\sum\limits_{m\geq 1}k_{0m}\rho_U^m}{1+\sum\limits_{n\geq 1}k_{n0}\rho_W^n+\sum\limits_{m\geq 1}k_{0m}\rho_U^m+\sum\sum\limits_{n\geq 1,m\geq 1}k_{nm}\rho_w^n\rho_U^m} \tag{7.9}$$

In our notation, if there are two subscripts then first will always correspond to water and the second to the osmolyte, ρ_w is the concentration of water in the aqueous solution of osmolyte, ρ_U is the concentration of co-solute in the aqueous solution of co-solute, $x(W=n,U=m)$ is the probability of observing a cluster that has n water and m osmolyte molecules, x_{*0} is the probability of observing a cavity free of osmolyte and x_{0*} is the probability of observing a cavity free of co-solute.

Now, we can compare the probability of observing an empty inner shell in pure water and in aqueous solution of a co-solute. We write:

$$\frac{x_0}{x_{00}}=\frac{1+\sum\limits_{n\geq 1\ ,m=0}k_{n0}\rho_w^n+\sum\limits_{n\geq 1}\sum\limits_{m\geq 1}k_{nm}\rho_w^n\rho_U^m}{1+\sum\limits_{n\geq 1}k_n\rho^n} \tag{7.10}$$

$$\frac{x_0}{x_{00}} = \frac{1 + \sum\limits_{n\geq 1\ ,m=0} k_{n0}\rho_w^n}{1 + \sum\limits_{n\geq 1} k_n\rho^n} \times \frac{1 + \sum\limits_{n\geq 1}\sum\limits_{m\geq 1} k_{nm}\rho_w^n\rho_U^m}{1 + \sum\limits_{n\geq 1\ ,m=0} k_{n0}\rho_w^n} \tag{7.11}$$

$$\frac{x_0}{x_{00}} = \frac{1 + \sum\limits_{n\geq 1\ ,m=0} \frac{k_{n0}\rho_w^n}{k_n\rho^n} k_n\rho^n}{1 + \sum\limits_{n\geq 1} k_n\rho^n} \times \frac{1 + \sum\limits_{n\geq 1}\sum\limits_{m\geq 1} k_{n0}\rho_w^n\ \ \frac{k_{nm}}{k_{n0}}\rho_U^m}{1 + \sum\limits_{n\geq 1\ ,m=0} k_{n0}\rho_w^n} \tag{7.12}$$

$$\frac{x_0}{x_{00}} = \Big(\sum_{n\geq 0\ ,m=0} \underline{x}(W=n)\ \frac{k_{n0}\rho_w^n}{k_n\rho^n}\Big) \times \frac{1 + \sum\limits_{n\geq 1}\sum\limits_{m\geq 1} k_{n0}\rho_w^n\ \ \frac{k_{nm}}{k_{n0}}\rho_U^m}{1 + \sum\limits_{n\geq 1\ ,m=0} k_{n0}\rho_w^n} \tag{7.13}$$

Notice the physical meaning of the first term on the right hand side. This is the dilution contribution. Physically it is related to the transfer of a n water cluster from water to osmolyte solution. Let's denote this by x^{dilution}.

$$\frac{x_0}{x_{00}} = x^{\text{dilution}} \times \Big(x(W=0|U=0) + \sum_{n\geq 1}\sum_{m\geq 1} x(W=n|U=0)\ \ \frac{k_{nm}}{k_{n0}}\rho_U^m\Big) \tag{7.14}$$

Notice the second term on the right hand side. This this is the related to the free energy of inserting cosolute molecules in a n water cluster. Let's denote this by x^{insert} Then

$$\frac{x_0}{x_{00}} = x^{\text{dilution}} \times x^{\text{insert}} \tag{7.15}$$

All these quantities can be calculated using our regularization scheme (as a test, we have already calculated them for a few cases). Notice that we have finally parsed the chemistry contribution to transfer into physically imaginable quantities.

8

Appendix

8.1 Enthalpy and entropy of solvation

Here we will derive an expression for entropy of solvation.

Let G(N,P,T) be the Gibbs free energy of a system of N particles at temperature T and pressure P. We introduce a solute particle in the system and the new Gibbs free energy is G(N+1,P,T). This solute particle is allowed to sample the entire volume of the system. We will denote enthalpy of the system by H and internal energy by U.

We write Gibbs-Helmholtz equation

$$\Big(\frac{\partial\big(\frac{G(N,P,T)}{T}\big)}{\partial T}\Big)_P = -\frac{H(N,P,T)}{T^2} \tag{8.1}$$

$$\Big(\frac{\partial\big(\frac{G(N+1,P,T)}{T}\big)}{\partial T}\Big)_P = -\frac{H(N+1,P,T)}{T^2} \tag{8.2}$$

Subtract the former from latter,

$$\Big(\frac{\partial\big(\frac{\Delta G(P,T)}{T}\big)}{\partial T}\Big)_P = -\frac{\Delta H(P,T)}{T^2} \tag{8.3}$$

Here $\Delta G(P,T)$ is $\frac{G(N+1,P,T)-G(N,P,T)}{1}$. The one in the denominator is the number of solute particles. So, $\Delta G(P,T)$ is chemical potential ($\mu = \mu^{id} + \mu^{ex}$) of solute.

μ^{id} is the ideal contribution to chemical potential ($k_BT \log \Lambda^3/V$) . μ^{ex} is the excess contribution to the chemical potential. Write H as U+PV.

$$k_B\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P + \Big(\frac{\partial\Big(\frac{\mu^{ex}}{T}\Big)}{\partial\ T}\Big)_P = -\frac{\Delta U(P,T)}{T^2} - P\frac{\Delta V(P,T)}{T^2} \tag{8.4}$$

ΔU is E^{ex}, defined as the difference in the internal energy of a solute-solvent system and just the solvent alone, and the kinetic energy of the solute particle ($\frac{3k_BT}{2}$). ΔV is the derivative of V w.r.t. number of solute particles at constant T and P. This is the partial molar volume of solute particles $\bar{V}$.

$$k_B\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P + \Big(\frac{\partial\Big(\frac{\mu^{ex}}{T}\Big)}{\partial\ T}\Big)_P = -\frac{E^{ex} + \frac{3k_BT}{2}}{T^2} - \frac{P\bar{V}}{T^2} \tag{8.5}$$

$$k_B\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P + \frac{1}{T}\Big(\frac{\partial\ \mu^{ex}}{\partial\ T}\Big)_P - \frac{\mu^{ex}}{T^2} = -\frac{E^{ex} + \frac{3k_BT}{2}}{T^2} - \frac{P\bar{V}}{T^2} \tag{8.6}$$

Multiply by T^2 and also identify entropy in L.H.S.

$$k_BT^2\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P - TS^{ex} - \mu^{ex} = -E^{ex} - \frac{3k_BT}{2} - P\bar{V} \tag{8.7}$$

$$-k_BT^2\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P + TS^{ex} = E^{ex} + \frac{3k_BT}{2} + P\bar{V} - \mu^{ex} \tag{8.8}$$

With the definition of partial molar volume

$$-k_BT^2\Big(\frac{\partial\ \log\frac{\Lambda^3}{V}}{\partial\ T}\Big)_P + TS^{ex} = E^{ex} + P\Big(k_BT\ k_T + V^{ex}\Big) - \mu^{ex} \tag{8.9}$$

$$-k_BT^2\Big(\frac{\partial\ \log\Lambda^3}{\partial\ T}\Big)_P + k_BT^2\Big(\frac{\partial\ \log V}{\partial\ T}\Big)_P + TS^{ex} = E^{ex} + \frac{3k_BT}{2} + P\Big(k_BT\ k_T + V^{ex}\Big) - \mu^{ex} \tag{8.10}$$

$$-k_BT^2\Big(\frac{\partial\ \log\Lambda^3}{\partial\ T}\Big)_P + \frac{k_BT^2}{V}\Big(\frac{\partial\ V}{\partial\ T}\Big)_P + TS^{ex} = E^{ex} + \frac{3k_BT}{2} + P\Big(k_BT\ k_T + V^{ex}\Big) - \mu^{ex} \tag{8.11}$$

identify thermal expansivity

$$-k_BT^2\Big(\frac{\partial\ \log\Lambda^3}{\partial\ T}\Big)_P+k_BT^2\alpha_P+TS^{ex}=E^{ex}+\frac{3k_BT}{2}+P\Big(k_BT\ k_T+V^{ex}\Big)-\mu^{ex} \quad (8.12)$$

$$-\frac{3k_BT^2}{\Lambda}\Big(\frac{\partial\ \Lambda}{\partial\ T}\Big)_P+k_BT^2\alpha_P+TS^{ex}=E^{ex}+\frac{3k_BT}{2}+P\Big(k_BT\ k_T+V^{ex}\Big)-\mu^{ex} \quad (8.13)$$

Because Λ is $\sqrt{\frac{h^2}{2\pi mk_BT}}$

$$-\frac{3k_BT^2}{\Lambda}\Big(\frac{\partial\ \Lambda}{\partial\ T}\Big)_P+k_BT^2\alpha_P+TS^{ex}=E^{ex}+\frac{3k_BT}{2}+P\Big(k_BT\ k_T+V^{ex}\Big)-\mu^{ex} \quad (8.14)$$

$$\frac{3k_BT}{2}+k_BT^2\alpha_P+TS^{ex}=E^{ex}+\frac{3k_BT}{2}+P\Big(k_BT\ k_T+V^{ex}\Big)-\mu^{ex} \quad (8.15)$$

$$TS^{ex}=E^{ex}-k_BT^2\alpha_P+P\Big(k_BT\ k_T+V^{ex}\Big)-\mu^{ex} \quad (8.16)$$

where α_p is the thermal expansivity, κ_T is the isothermal compressibility and V^{ex} is the excess volume. If these three are negligible then

$$Ts^{\mathrm{ex}} \approx E_{sw}+E_{reorg}-\mu^{\mathrm{ex}} \quad (8.17)$$

And the excess enthalpy of hydration

$$h^{\mathrm{ex}}=E^{\mathrm{ex}}=E_{sw}+E_{reorg} \quad (8.18)$$

E^{ex}, is the sum of solute-solvent binding energy (E_{sw}) and reorganization energy (E_{reorg}).

Bibliography

[1] C. Tanford, and J. Reynolds, *Nature's Robots: A History of Proteins*, Oxford University Press, USA, (2004).

[2] R. E. Dickerson, and I. Geis, *The structure and action of proteins*, W. A. Benjamin, Inc., (1969).

[3] M. Williamson *How proteins work*, Garland Science, Taylor and Francis Group, LLC, (2012).

[4] C. K. Mathews, and K. E. van Holde, *Biochemistry*, The Benjamin/Cummings Publishing Company, (1990).

[5] H. J. Dyson, and P. E. Wright, Nature Reviews Molecular Cell Biology **6**, 197 (2005).

[6] C. M. Dobson, Nature **426**, 884 (2003).

[7] J. W. Yewdell, E. Reits, and J. Neifjes, Nature Reviews Immunology **3**, 952 (203).

[8] C. B. Anfinsen, Science **181**, 4096 (1973).

[9] A. Ben-Naim, Journal of Biomolecular Structure and Dynamics **30**, 113 (2012).

[10] G. Kolata, Science **233**, 233 (1986).

BIBLIOGRAPHY

[11] C. Tanford, *The hydrophobic effect: formation of micelles and biological membranes*, Johns Willey and sons, New York, (1980).

[12] M. Strong, M. R. Sawaya, S. Wang, M. Phillips, D. Cascio, and D. Eisenberg, Proceedings of National Academy of Sciences **103**, 8060 (2006).

[13] J. T. Edsall, Proceedings of the American Philosophical Society **129**, 371 (1985).

[14] T. Svedberg, ..., J. F. Danielli, ..., Proceedings of the Royal Society of London **170**, 73 (1939).

[15] I. Langmuir, and D. Wrinch, Nature **143**, 49 (1939).

[16] H. S. Frank, and M. W. Evans, Journal of Chemical Physics **13**, 507 (1945).

[17] W. Kauzmann, Advances in Protein Chemistry **14**, 1 (1959).

[18] W. Kauzmann, Nature **325**, 763 (1987).

[19] C. Tanford, Advances in Protein Chemistry **24**, 1 (1970).

[20] K. A. Dill, and J. L. MacCallum, Science **338**, 1042 (2012).

[21] A. Ben-Naim, Journal of Chemical Physics **135**, 085104 (2012).

[22] J. N. Onuchic, Z.Luthey-Schulten, and P. G. Wolynes, Annual Reviews of Physical Chemistry **48**, 545 (1997).

[23] K. A. Dill, and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*, Garland Science, (2010).

[24] H. Bohr, and S. Brunak, *Protein Folds: A Distance-Based Approach*, CRC Press, (1995).

BIBLIOGRAPHY

[25] K. Huang, *Lectures on Statistical Physics and Protein Folding*, World Scientific Pub. Co. Inc., (2005).

[26] V. Muñoz, *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*, RSC Publishing, (2008).

[27] D. Voet, and J. G. Voet, *Biochemistry*, John Willey and Sons, INC, (2011).

[28] A. Kessel, and N. Ben-Tal, *Introduction to proteins: structure function, and motion*, CRC Press, (2011).

[29] A. M. Lesk, *Introduction to protein science: architecture, function and genomics*, Oxford Univeristy Press, (2010).

[30] D. T. Haynie, *Biological thermodynamics*, Cambridge Univeristy Press, (2001).

[31] A. Karshikoff, *Non-covalent interactions in proteins*, Imperial College Press, (2006).

[32] K. A. Dill, Biochemistry **29**, 7133 (1990).

[33] I. M. Klotz, and J. S.Frenzen, Jouranl of American Chemical Society **84**, 3461 (1962).

[34] A. R. Fersht, Trends in Biochemical Sciences **12**, 301 (1987).

[35] A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, W. H. Freeman and Compnay, (1999).

[36] A. Ben-Naim, Jounal of Physical Chemistry **95**, 1437 (1991).

[37] Q. Zou, B. J. Bennion, V. Daggett, and K. P. Murphy, Journal of American Chemical Soceity **124**, 1192 (2002).

[38] B. J. Bennion, and V. Daggett, Proceedings of National Academy of Sciences **101**, 6433 (2004).

[39] C. Y. Hu, G. C. Lynch, H.Kokubo, and B. M. Pettitt, Proteins: Structure, Function, and Bioinformatics **78**, 695 (2010).

[40] D. Canchi, D. Paschek, and A. E. Garcia, Journal of American Chemical Society **132**, 2338 (2010).

[41] A. Yang, and B. Honig, Journal of Molecular Biology **252**, 351 (1995).

[42] F. Avbelj, and L. Fele, Journal of Molecular Biology **279**, 665 (1998).

[43] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, Proceedings of National Academy of Sciences **103**, 16623 (2006).

[44] T. O. Street, B. W. Bolen, and G. D. Rose, Proceedings of National Academy of Sciences **103**, 13997 (2006).

[45] A. Ben-Naim, International Journal of Physics **1**, 66 (2013).

[46] D. P. Teufel, C. M. Johnson, J. K. Lum, and H. Neuweiler, Journal of Molecular Biology **409**, 250 (2011).

[47] R. L. Baldwin, and G. D. Rose, Trends in Biochemical Sciences **24**, 26 (1999).

[48] R. L. Baldwin, and G. D. Rose, Trends in Biochemical Sciences **24**, 77 (1999).

[49] M. Karplus, and D. L.Weaver, Protein Science **3**, 650 (1994).

[50] G. R. Bowman, V. A. Voelz, and V. S. Pande, Journal of American Chemical Society **133**, 664 (2011).

[51] H. Gong, L. L. Porter, and G. D. Rose, Protein Science **20**, 417 (2011).

[52] J. Chen, and C. L. Brooks III, Physical Chemistry Chemical Physics **10**, 471 (2008).

[53] C. Y. Hu, L. C. Lynch, H. Kokubo, and B. M. Pettitt, Proteins: Structure, Function, and Bioinformatics **78**, 695 (2010).

[54] H. T. Tran, A. Mao, and R. V. Pappu, Jouranl of American Chemical Society **130**, 7380 (2008).

[55] B. Widom, Journal of Chemical Physics **39**, 2808 (1963).

[56] L. R. Pratt, Annual Reviews of Phyical Chemistry **53**, 409 (2002).

[57] M. E. Paulaitis and L. R. Pratt, Advances in Protein Chemistry **62**, 283 (2002).

[58] S. Merchant, *Regularizing free energy calculations to study ion specific effects in biology*, Johns Hopkins University, 2011.

[59] C. Chipot, and A. Pohorille, Chapter 9 in *Free energy calculations: theory and applications in chemistry and biology*, Springer, (2007).

[60] T. L. Beck, M. Paulaitis, and L. R. Pratt, *The potential distribution theorem and models of molecular solutions*, Cambridge University Press, (2006).

[61] L. R. Pratt and A. Pohorille, Proceedings of National Academy of Sciences **89**, 2995 (1992).

[62] V. Weber and D. Asthagiri, Journal of Chemical Theory and Computation **8**, 3409 (2012).

[63] V. Weber, S. Merchant, and D. Asthagiri, Journal of Chemical Physics **135**, 181101 (2011).

[64] V. Weber, and D. Asthagiri, Journal of Chemical Physics **133**, 141101 (2010).

[65] S. Merchant, J. K. Shah, and D. Asthagiri, Journal of Chemical Physics **134**, 124514 (2010).

[66] S. Merchant, P. D. Dixit, K. R. Dean, and D. Asthagiri, Journal of Chemical Physics **135**, 054505 (2011).

[67] S. Merchant, and D. Asthagiri, Journal of Chemical Physics **130**, 195102 (2009).

[68] D. Asthagiri, H. S. Ashbaugh, A. Piryatinski, M. E. Paulaitis, and L. R. Pratt, Journal of American Chemical Society **129**, 10133 (2007).

[69] J. K. Shah, D. Asthagiri, L. R. Pratt, and M. E. Paulaitis, Journal of Chemical Physics **127**, 144508 (2007).

[70] S. Utiramerur and M. E. Paulaitis, Journal of Chemical Physics **132**, 155102 (2010).

[71] D. Asthagiri, S.Merchant, and L. R. Pratt, Journal of Chemical Physics **128**, 244512 (2008).

[72] D. S. Tomar, V. Weber, and D. Asthagiri, Biophysical Journal **105**, 1482 (2013).

[73] D. S. Tomar, V. Weber, B. M Pettitt, and D. Asthagiri, Journal of Physical Chemistry B **118**, 4080 (2014).

[74] M. Auton and D. W. Bolen, Biochemistry **43**, 1329 (2004).

[75] R. F. Greene and C. N. Pace, Journal of Biological Chemistry **249**, 5388 (1974).

[76] D. R. Canchi, P. Jayasimha, D. C. Rau, G. I. Makhatadze, and A. E.García, Journal of Physical Chemistry B **116**, 12095 (2012).

[77] S. Weerasinghe and P. E. Smith, Journal of Physical Chemistry B **107**, 3891 (2003).

[78] S. Weerasinghe and P. E. Smith, Journal of Chemical Physics **121**, 2180 (2004).

[79] K. M. Kast, J. Brickmann, S. M. Kast, and R. S. Berry, Journal of Physical Chemistry B **107**, 5342 (2003).

[80] R. Godawat, S. N. Jamadagni, and S. Garde, Journal of Physical Chemistry B **114**, 2246 (2010).

[81] S. N. Timasheff, Proceedings of National Academy of Sciences **99**, 9721 (2002).

[82] J. A. Schellman, Biophysical Chemistry **96**, 91 (2002).

[83] P. H. Yancey, W. R. Blake, and J. Conley, Comparative Biochemistry and Physiology Part A **133**, 667 (2002).

[84] G. Konig, S. Bruckner, and S. Boresch, Biophysical Journal **104**, 453 (2013).

[85] R. B. Simpson, and W.Kauzmann, Journal of American Chemical Society **75**, 5139 (1953).

[86] D. O. Alonso, K. A.Dill, and Biochemistry **30**, 5974 (1991).

[87] I. T. Li, and G. C. Walker, Journal of American Chemical Society **132**, 6530 (2010).

[88] I. T. Li, and G. C. Walker, Proceedings of National Academy of Sciences **108**, 16527 (2011).

[89] B. J. Bennion, and V. Daggett, Proceedings of National Academy of Sciences **100**, 5142 (2003).

[90] D. Tobi, R. Elber, and D. Thirumalai, Biopolymers **68**, 359 (2003).

[91] A. Wallqvist, D. G. Covell, and D. Thirumalai Journal of American Chemical Society **120**, 427 (1998).

[92] L. Hua, R. Zhou, D. Thirumalai, and B. J. Berne Proceedings of National Academy of Sciences **104**, 16928 (2008).

[93] D. W. Bolen, and G. D. Rose, Annual Reviews of Biochemistry **77**, 339 (2008).

[94] D. W. Bolen, and I. V. Baskakov, Journal of Molecular Biology **310**, 955 (2001).

[95] P. L. Privalov, and N. N. Khechinashvili, Journal of Molecular Biology **86**, 665 (1974).

[96] D. M. Huang, and D. Chandler, Proceedings of National Academy of Sciences **97**, 8324 (2000).

[97] B. Moeser, and D. Horinek, Journal of Physical Chemistry B **118**, 107 (2014).

[98] H. Kokubo, C. Y. Hu, and B. M. Pettitt, Journal of American Chemical Society **133**, 1849 (2011).

[99] H. Kokubo, R. C. Harris, D. Asthagiri, and B. M. Pettitt, Journal of Physical Chemistry B **117**, 16428 (2013).

[100] R. Zangi, R. Zhou, and B. J. Berne, Journal of American Chemical Society **131**, 1535 (2009).

[101] A. Ben-Naim, Open Journal of Biophysics **1**, 1 (2011).

[102] P. L. Privalov, Advances in Protein Chemistry **33**, 167 (1979).

[103] R. L. Baldwin, Proceedings of National Academy of Sciences **83**, 8069 (1986).

[104] R. L. Baldwin, Biophysical Chemistry **101-102**, 203 (2002).

[105] R. L. Baldwin, Journal of Biological Chemistry **278**, 17581 (2003).

[106] P. L. Privalov, Critical Reviews in Biochemistry and Molecular Biology **25**, 281 (1990).

[107] S. Garde, G. Hummer, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, Physical Review Letters **77**, 4966 (1996).

[108] S. Cabani, and P. Gianni, Journal of the Chemical Society, Faraday Transactions 1 **75**, 1184 (1979).

[109] A. Ben-Naim, and Y. Marcus, Journal of Chemical Physics **81**, 2016 (1984).

[110] W. Jorgensen, J. Chandrasekhar, ..., and M. L. Klein, Journal of Chemical Physics **79**, 926 (1983).

[111] M. J. Frisch, ..., and D.J. Fox, Gaussian Inc. Wallingford CT 2009, 2009.

[112] E. Neria, S. Fischer, and M. Karplus, Journal of Chemical Physics **105**, 1902 (1996).

[113] A. D. MacKerell Jr., M. Feig, and C. L. Brooks III, Journal of Computational Chemistry **25**, 1400 (2004).

[114] A. D. MacKerell Jr., D. Bashford, ..., and M. Karplus, Journal of Physical Chemistry B **102**, 3586 (1998).

[115] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, Journal of Physical Chemistry **91**, 6269 (1987).

[116] G. I. Makhatadze and P. L. Privalov, Advances in Protein Chemistry **47**, 307 (1995).

[117] J. M. Richardson and G. I. Makhatadze, Journal of Molecular Biology **335**, 1029 (2004).

[118] A. García and K. Y. Sanbonmatsu, Proceedings of National Academy of Sciences **99**, 2782 (2002).

[119] R. B. Best and G. Hummer, Journal of Physical Chemistry B **113**, 9004 (2009).

[120] R. L. Baldwin, Journal of Molecular Biology **371**, 283 (2007).

[121] A. Chakrabartty and R. L. Baldwin, Advances in Protein Chemistry **46**, 141 (1995).

[122] R. Staritzbichler, W. Gu, and V. Helms, Journal of Physical Chemistry B **109**, 19000 (2005).

[123] G. König and S. Boresch, Journal of Physical Chemistry B **113**, 8967 (2009).

[124] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, Protein Science **4**, 561 (1995).

[125] E. Darve, D. Rodriguez-Gómez, and A. Pohorille, Journal Chemical Physics **128**, 144120 (2008).

[126] J. Hénin, G. Fiorin, C. Chipot, and M. L. Klein, Journal of Chemical Theory and Computation **6**, 35 (2010).

[127] F. H. Stillinger, Jornal of Solution Chemistry **2**, 141 (1973).

[128] D. Chandler, Nature **437**, 640 (2005).

[129] H. S. Ashbaugh and L. R. Pratt, Reviews of Modern Physics **78**, 159 (2006).

[130] A. L. Ferguson, P. G. Debenedetti, and A. Z. Panagiotopoulos, J. Phys. Chem. B **113**, 6405 (2009).

[131] M. V. Athawale, G. Goel, T. Ghosh, T. M. Truskett, and S. Garde, Proceedings of National Academy of Sciences **104**, 733 (2007).

[132] S. Gnanakaran and A. E. García, Proteins **59**, 773 (2005).

[133] E. J. Sorin and V. S. Pande, Biophysical Journal **88**, 2472 (2005).

[134] T. Ooi and M. Oobatake, Proceedings of National Academy of Sciences **88**, 2859 (1991).

[135] S. Marqusee, V. H. Robbins, and R. L. Baldwin, Proceedings of National Academy of Sciences **86**, 5286 (1989).

[136] J. M. Scholtz, S. Marqusee, R. L. Baldwin, E. J. York, J. M. Stewart, M. Santoro, and D. W. Bolen, Proceedings of National Academy of Sciences **88**, 2854 (1991).

[137] E. J. Spek, C. A. Olson, Z. S. Shi, and N. R. Kallenbach, Journal of American Chemical Society **121**, 5571 (1999).

[138] Z. Shi, K. Chen, Z. Liu, and N. R. Kallenbach, Chemical Reviews **106**, 1877 (2006).

[139] J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, Journal of American Chemical Society **129**, 1179 (2007).

[140] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, Journal of American Chemical Society **112**, 6127 (1990).

[141] J. Weiser, P. Senkin, and W. C. Still, Journal of Computational Chemistry **20**, 217 (1999).

[142] A. Onufriev, D. Bashford, and D. A. Case, Journal of Physical Chemistry B **104**, 3712 (2000).

[143] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, Jr., Journal of Chemical Theory and Computation **8**, 3257 (2012).

[144] R. B. Best, N. V. Buchete, and G. Hummer, Biophysical Journal **95**, L07 (2008).

[145] A. D. MacKerell, Jr., Journal of Computational Chemistry **25**, 1584 (2004).

[146] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, Annual Reviews of Biophysics **37**, 289 (2008).

[147] C. Chothia, Nature **254**, 304 (1974).

[148] C. Tanford, Journal of American Chemical Society **84**, 4240 (1962).

[149] A. D. Robertson, and K. P. Murphy, Chemical Reviews **97**, 1251 (1997).

[150] W. V. Kayser, Journal of Colloid and Interface Science **56**, 622 (1976).

[151] N. B. Vargaftik, B. N. Volkov, and L. D. Voljak, Journal of Physical and Chemical Reference Data **12**, 817 (1983).

[152] M. Auton, J. Rösgen, ..., and D. W. Bolen, Biophysical Chemistry **159**, 90 (2011).

[153] F. Avbelj and R. L. Baldwin, Proteins: Structure, Function, and Bioinformatics **63**, 283 (2006).

[154] F. Avbelj and R. L. Baldwin, Proceedings of National Academy Sciences **106**, 3137 (2009).

[155] B. Lee and F. M. Richards, Journal Of Molecular Biology **55**, 379 (1971).

[156] A. Bondi Journal of physical chemistry **68**, 441 (1964).

[157] M. Sanner, A. J. Olson, and J. C. Spehner Biopolymers **38**, 305 (1996).

[158] Y. Nozaki and C. Tanford, Journal of American Chemical Society **238**, 4074 (1963).

[159] M. A. Roseman Journal of Molecular Biology **200**, 513 (1988).

[160] J. L. Cornette, ...,and C. DeLisi Journal of Molecular Biology **195**, 659 (1987).

[161] G. Hummer, and S. Attila Journal of Chemical Physics **105**, 2004 (1996).

[162] P. A. Alexander, Y. He, ..., and P.N. Bryan Proceedings of National Academy of Sciences **106**, 21149 (2009).

Vita

Dheeraj Singh Tomar was born in Dholpur (Rajasthan), India on January 1st, 1987, to Sunita and Umacharan Tomar as the eldest of their three children. He did his secondary schooling at the Government Senior Secondary School, Saipau (Dholpur), and senior secondary schooling at A.G.M. school, Talwandi, Kota (Rajasthan), India. He finished his Bachelor of Technology in Chemical Engineering in 2009 at Indian Institute of Technology Guwahati, India. After that, he worked on computer simulation of adsorption of gases for a brief period at Indian Institute of Technology Guwahati. In 2010, he joined the Johns Hopkins University as a PhD student in the department of Chemical and Biomolecular Engineering. In last four years at JHU, he has worked on solvent effects in protein solution thermodynamics.

He will join Pfizer Inc. at St. Louis, Missori in August 2014 as a Postdoctoral Fellow.